



TUGAS AKHIR - SS 141501

**KLASIFIKASI *DYNAMIC FINANCIAL DISTRESS*
PERUSAHAAN MANUFAKTUR YANG TERDAFTAR
DI BURSA EFEK INDONESIA TAHUN 2012-2014
MENGUNAKAN REGRESI LOGISTIK BINER
DAN *CLASSIFICATION ANALYSIS & REGRESSION TREE***

**MUNIROH
NRP 1314 105 021**

**Dosen Pembimbing
Dr. Drs. Agus Suharsono, MS**

**PROGRAM STUDI S1
JURUSAN STATISTIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2016**



FINAL PROJECT - SS 141501

**CLASSIFICATION OF DYNAMIC FINANCIAL
DISTRESS FOR MANUFACTURING COMPANY
THAT LISTED IN INDONESIA STOCK EXCHANGE
2012-2014 USING BINARY LOGISTIC REGRESSION
AND CLASSIFICATION ANALYSIS & REGRESSION TREE**

**MUNIROH
NRP 1314 105 021**

**Supervisor
Dr. Drs. Agus Suharsono, MS**

**UNDERGRADUATE PROGRAMME
DEPARTMENT OF STATISTICS
FACULTY OF MATHEMATICS AND NATURAL SCIENCES
INSTITUT TEKNOLOGI SEPULUH NOPEMBER
SURABAYA 2016**

LEMBAR PENGESAHAN

**KLASIFIKASI *DYNAMIC FINANCIAL DISTRESS*
PERUSAHAAN MANUFAKTUR YANG TERDAFTAR
DI BURSA EFEK INDONESIA TAHUN 2012-2014
MENGUNAKAN REGRESI LOGISTIK BINER
DAN *CLASSIFICATION ANALYSIS & REGRESSION
TREE***

TUGAS AKHIR

**Diajukan Untuk Memenuhi Salah Satu Syarat
Memperoleh Gelar Sarjana Sains
pada
Program Studi S-1 Jurusan Statistika
Fakultas Matematika dan Ilmu Pengetahuan Alam
Institut Teknologi Sepuluh Nopember Surabaya**

Oleh :

MUNIROH

NRP.1314 105 021

Disetujui oleh Pembimbing Tugas Akhir :

Dr. Drs. Agus Suharsono, MS

NIP. 19580823 198403 1 003

(.....)

Mengetahui,

Ketua Jurusan Statistika FMIPA-ITS

Dr. Suhartono

NIP. 19710929 199512 1 001

SURABAYA, JULI 2016

**Klasifikasi *Dynamic Financial Distress* Perusahaan Manufaktur
yang Terdaftar di Bursa Efek Indonesia Tahun 2012-2014
Menggunakan Regresi Logistik Biner dan *Classification Analysis &
Regression Tree***

Nama Mahasiswa : Muniroh
NRP : 1314 105 021
Jurusan : Statistika FMIPA-ITS
Pembimbing : Dr. Drs. Agus Suharsono, MS

Abstrak

Krisis keuangan global tak hanya berdampak pada sektor riil, tapi juga sangat memukul sektor finansial. Sektor yang terkena imbasan krisis keuangan global adalah seluruh sektor bidang kehidupan. Namun yang paling tampak gejalanya adalah sektor bidang industri manufaktur. Beberapa perusahaan manufaktur yang mengalami masalah keuangan mencoba mengatasi masalah tersebut dengan melakukan pinjaman dan penggabungan usaha, atau sebaliknya ada yang menutup usahanya. Tujuan dilakukan penelitian ini yaitu untuk mengetahui faktor-faktor yang diduga berpengaruh signifikan terhadap financial distress, Mengetahui model klasifikasi financial distress. Sehingga, pada tugas akhir ini akan dicoba melakukan klasifikasi dynamic financial distress perusahaan manufaktur yang terdaftar di bursa efek indonesia tahun 2012-2014 menggunakan regresi logistik biner dan classification analysis & regression tree (CART). Sebelum menganalisis dengan metode regresi logistik biner dan CART data terlebih dahulu di preprocessing imbalance data agar seimbang. Berdasarkan hasil penelitian dari regresi logistik biner terdapat dua variabel prediktor yang berpengaruh signifikan terhadap kondisi finansial distress perusahaan manufaktur yaitu variabel rasio likuiditas dan rasio aktivitas, dengan ketepatan klasifikasi sebesar 84,8%. Metode CART menghasilkan pohon klasifikasi maksimum sama dengan pohon klasifikasi optimum, dengan simpul utamanya adalah variabel rasio solvabilitas. Nilai ketepatan klasifikasi sebesar 96,3%. Perbandingan kedua metode menghasilkan bahwa metode untuk data training CART lebih besar 11,5%, tetapi untuk data testing regresi logistik biner unggul 2% dari metode CART.

Kata kunci: *CART, Financial Distress, Klasifikasi Perusahaan Manufaktur, Regresi Logistik Biner.*

**Classification Of Dynamic Financial Distress Manufacturing
Company That Listed In Indonesia Stock Exchange 2012-2014
Using Binary Logistic Regression And Classification Analysis &
Regression Tree**

Name : Muniroh
NRP : 1314 105 021
Department : Statistika FMIPA-ITS
Supervisor : Dr. Drs. Agus Suharsono, MS

Abstract

Financial Global Crisis does not only have an impact on the real sector, but also hit the financial sector. The sectors that affected by the contagious nature of the global financial crisis is the whole sectors of life. However, the most visible symptoms are the sectors of manufacturing industry. Some manufacturing companies that are experiencing financial problems trying to overcome these problems by borrowing and merger, or otherwise there were shut down. The purpose of this study is to determine the factors that allegedly significantly influence financial distress, and knowing classification model of financial distress. Thus, this final project will try to classify the dynamic financial distress companies that listed on stock exchanges in Indonesia 2012-2014 using binary logistic regression and classification and regression tree analysis (CART). Before analyzing the binary logistic regression and CART, first the data must be balanced using SMOTE. According to the research for binary logistic regression there are two predictor variables that significantly influence for financial condition of distress for manufacturing company that variable is liquidity ratio and the activity ratio, a classification accuracy is 84,8%. CART classification tree method produces that maximum classification tree is equal to the optimum classification tree, with the primary node is variable solvency ratio, and the value of a classification accuracy is 96,3%. Comparison of two methods produce that for training data CART has a value greater 11.5% classification accuracy than binary logistic regression method but for testing data, binary logistic regression has a value greater classification accuracy 2% than CART.

Keywords : CART, Financial Distress, Classification Manufacturing Company, Binary Logistic Regression.

DAFTAR ISI

	Halaman
HALAMAN JUDUL	
LEMBAR PENGESAHAN	
ABSTRAK	i
ABSTRACT	iii
KATA PENGANTAR	v
DAFTAR ISI	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL	xi
DAFTAR LAMPIRAN	xv
 BAB I PENDAHULUAN	
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	4
1.3 Tujuan Penelitian	5
1.4 Manfaat Hasil Penelitian.....	5
1.5 Batasan Masalah	5
 BAB II TINJAUAN PUSTAKA	
2.1 SMOTE.....	7
2.2 Regresi Logistik Biner.....	8
2.2.1 Estimasi Parameter	9
2.2.2 Pengujian Parameter	10
2.3 <i>Classification Analysis and Regression Tree (CART)</i>	11
2.4 <i>Financial Distress</i>	19
2.5 Rasio Keuangan Model Altman.....	20
 BAB III METODOLOGI PENELITIAN	
3.1 Sumber Data.....	23
3.2 Variabel Penelitian	24
3.3 Langkah Analisis.....	24

BAB IV HASIL DAN PEMBAHASAN

4.1	Deskriptif Data Perusahaan Manufaktur	27
4.2	<i>Preprocessing Imbalanced Data</i>	33
4.3	Regresi Logistik Biner Data <i>Training</i> 2012-2013	37
4.3.1	Deteksi Multikolinearitas.....	37
4.3.2	Pengujian Parameter	39
4.3.3	Ketepatan Klasifikasi	40
4.4	Klasifikasi <i>Financial Distress</i> dengan Metode Classification Analysis and Regression Tree (CART)	41
4.4.1	Pembentukan Pohon Klasifikasi	42
4.4.2	Penentuan Pohon Klasifikasi Optimum.....	44
4.4.3	Ketepatan Klasifikasi Pohon CART	47
4.5	Perbandingan Hasil Klasifikasi Antara Metode Regresi Logistik Biner dan CART	51

BAB V KESIMPULAN DAN SARAN

5.1	Kesimpulan.....	53
5.2	Saran	54

DAFTAR PUSTAKA	55
-----------------------------	----

LAMPIRAN	57
-----------------------	----

DAFTAR TABEL

	Halaman
Tabel 2.1	Variabel Ilustrasi 14
Tabel 2.2	Data Sampel untuk Ilustrasi..... 14
Tabel 2.3	Rata-rata Dua Amatan Sampel Berurutan Variabel X_1 16
Tabel 2.4	Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> 0,055..... 16
Tabel 2.5	Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> 0,115..... 16
Tabel 2.6	Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> 0,145..... 16
Tabel 2.7	Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> 0,160..... 17
Tabel 2.8	Rata-rata Dua Amatan Sampel Berurutan Variabel X_2 17
Tabel 2.9	Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> 2,49..... 17
Tabel 2.10	Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> 2,63..... 18
Tabel 2.11	Perhitungan <i>Goodness of Split</i> untuk <i>Threshold</i> 2,74..... 18
Tabel 2.12	Perbandingan <i>Goodness of Split</i> dari Setiap Kemungkinan Pemilahan Satu 18
Tabel 3.1	Variabel Penelitian 24
Tabel 4.1	Persentase Perusahaan Manufaktur Mengalami <i>Financial Distress</i> dan <i>Non-Financial Distress</i> per Sektor Manufaktur Tahun 2013-2014 28
Tabel 4.2	Persentase Perusahaan Manufaktur Mengalami <i>Financial Distress</i> dan <i>Non-Financial Distress</i> per Sektor Manufaktur Tahun 2012-2013 28
Tabel 4.3	Statistika Deskriptif Variabel Prediktor per Masing-masing Kelas <i>Financial Distress</i> dan <i>Non-Financial Distress</i> Tahun 2013-2014..... 29

Tabel 4.4	Statistika Deskriptif Variabel Prediktor per Masing-masing Kelas <i>Financial Distress</i> dan <i>Non-Financial Distress</i> Tahun 2012-2013	31
Tabel 4.5	Distribusi Data Simulasi Sebelum dan Setelah SMOTE Tahun 2013-2014... ..	34
Tabel 4.6	Data Simulasi Setelah Menggunakan SMOTE Tahun 2013-2014.....	35
Tabel 4.7	Distribusi Data Simulasi Sebelum dan Setelah SMOTE Tahun 2012-2013	36
Tabel 4.8	Data Simulasi Setelah Menggunakan SMOTE Tahun 2012-2013.....	37
Tabel 4.9	Tabel Korelasi Antar Variabel Prediktor Tahun 2012-2013	38
Tabel 4.10	Nilai VIF Masing-masing Variabel Prediktor	38
Tabel 4.11	Uji Serentak Tahun 2012-2013.....	39
Tabel 4.12	Uji Parsial Tahun 2012-2013.....	40
Tabel 4.13	Ketepatan Klasifikasi Data <i>Training</i> Tahun 2012-2013	40
Tabel 4.14	Tabel Ketepatan Klasifikasi Data <i>Training</i> dan Data <i>Testing</i>	41
Tabel 4.15	Variabel Penting Pembentukan Pohon Klasifikasi Maksimum.....	42
Tabel 4.16	Urutan Pembentukan Pohon Klasifikasi (<i>Tree Sequence</i>).....	43
Tabel 4.17	Variabel Penting Pembentukan Pohon Klasifikasi Optimum.....	45
Tabel 4.18	Kelas Perusahaan Manufaktur pada Masing-masing Simpul Terminal.....	46
Tabel 4.19	Karakteristik Kelas Perusahaan Manufaktur Menurut Presentase Kelas Terbesar.....	47
Tabel 4.20	Klasifikasi Data <i>Training</i> pada Pohon Klasifikasi Optimum Sebelum SMOTE... ..	48
Tabel 4.21	Ketepatan Klasifikasi Data <i>Testing</i> Sebelum SMOTE.....	48

Tabel 4.22	Klasifikasi Data <i>Training</i> pada Pohon Klasifikasi Optimum Sesudah SMOTE.....	49
Tabel 4.23	Ketepatan Klasifikasi Data <i>Testing</i> pada Pohon Klasifikasi Optimum Sesudah SMOTE.....	50
Tabel 4.24	Perbandingan Ketepatan Klasifikasi Pohon Optimum Sebelum dan Sesudah SMOTE	51
Tabel 4.25	Perbandingan Ketepatan Klasifikasi Antara Metode Regresi Logistik dan Metode CART	52

BAB I

PENDAHULUAN

1.1 Latar Belakang

Krisis keuangan global tak hanya berdampak pada sektor riil, tapi juga sangat memukul sektor finansial. Sektor yang terkena imbasan krisis keuangan global adalah seluruh sektor bidang kehidupan. Namun yang paling tampak gejalanya adalah sektor bidang industri manufaktur. Di Amerika Serikat, krisis global telah mendorong sektor finansial merugi dengan hilangnya aset mereka hingga 50 triliun dollar AS. Keadaan Indonesia tidak beda jauh dengan Amerika, dimana krisis global telah menggerus laba BUMN dan menyebabkan konglomerat Indonesia kehilangan sekitar 60 persen kekayaannya, dan juga telah mengakibatkan kurang lebih 30.000 pekerja kehilangan pekerjaan maupun penghasilan. Keadaan industri manufaktur yang saat ini mengalami penurunan kinerja ekspor mengalami defisit neraca perdagangan pada tahun 2011-2013. Berdasarkan artikel Kementerian Perindustrian Republik Indonesia, nilai ekspor industri Manufaktur pada tahun 2011 sebesar USD 122,19 miliar, dan terjadi penurunan pada tahun 2012 sebesar 4,95%, dengan nilai ekspor sebesar USD 116,14. Penurunan nilai ekspor ini berbanding terbalik dengan nilai impor industri manufaktur yang semakin meningkat. Beberapa perusahaan manufaktur yang mengalami masalah keuangan mencoba mengatasi masalah tersebut dengan melakukan pinjaman dan penggabungan usaha, atau sebaliknya ada yang menutup usahanya.

Financial distress adalah suatu konsep luas yang terdiri dari beberapa situasi dimana suatu perusahaan menghadapi masalah kesulitan keuangan. Istilah umum untuk menggambarkan situasi tersebut adalah kebangkrutan, kegagalan, ketidak mampuan melunasi hutang, dan *default*. *Insolvency* dalam kebangkrutan menunjukkan kekayaan bersih negatif. Ketidakmampuan melunasi utang menunjukkan kinerja negatif dan menunjukkan

adanya masalah likuiditas. Kondisi ini pada umumnya ditandai antara lain dengan adanya dua penundaan pengiriman, kualitas produk yang menurun, dan penundaan pembayaran tagihan dari bank. Apabila kondisi *financial distress* ini diketahui, diharapkan dapat dilakukan tindakan untuk memperbaiki situasi tersebut sehingga perusahaan tidak akan masuk pada tahap kesulitan yang lebih berat seperti kebangkrutan ataupun likuidasi. Kebangkrutan suatu perusahaan dapat dilihat dan diukur melalui laporan keuangan. Laporan Keuangan yang diterbitkan oleh perusahaan merupakan salah satu sumber informasi mengenai posisi keuangan perusahaan, kinerja serta perubahan posisi keuangan perusahaan, yang sangat berguna untuk mendukung pengambilan keputusan yang tepat, data keuangan harus dikonversi menjadi informasi yang berguna dalam pengambilan keputusan ekonomis. Hal ini ditempuh dengan cara melakukan analisis dalam bentuk rasio-rasio keuangan (Ardhinta, 2009).

Menurut Sin & Li (2011), penelitian sebelumnya pada prediksi *financial distress* kebanyakan hanya berfokus pada model statis untuk memprediksi model, dimana model peramalannya dibentuk dari sampel data pada waktu tertentu saja. Seiring berjalannya waktu, model statis tidak dapat meramalkan kondisi *financial distress* dalam perubahan lingkungan ekonomi maupun perubahan operasional perusahaan dengan baik dan efektif. Karakteristik pada prediksi *financial distress* pada kenyataannya tidak pernah terpikirkan sebelumnya. Sehingga, untuk mengatasi lingkungan operasional perusahaan yang dinamis, sesuai perubahan waktu, penelitian pada model prediksi *financial distress* harus dilakukan dengan merubah asumsi-asumsi yang ada pada penelitian sebelumnya yaitu bahwa jumlah sampel data tidak berubah atau hanya pada periode waktu tertentu saja. Konsep *dynamic financial distress prediction* terdiri dari *instance selection*, *financial distress prediction modeling*, dan *future prediction*.

Pada dunia perbankan, indikasi awal terjadinya *financial distress* dapat diketahui dari laporan laba rugi, dimana bank

mengalami laba bersih negatif dan mengalami negatif *spread* akibat rendahnya biaya bunga pinjaman daripada bunga simpanan. Hofer (1980) dan Whitaker (1999) mengumpamakan kondisi *financial distress* sebagai kondisi dari perusahaan yang mengalami laba bersih (*net profit*) negatif selama beberapa tahun. Selain itu, penghapusan pencatatan saham dari Bursa akibat dari menurunnya kinerja juga merupakan indikasi awal perusahaan yang mengalami kebangkrutan (Hadi dan Anggraeni, 2008). Bursa Efek Indonesia (BEI), merupakan bursa hasil penggabungan dari Bursa Efek Jakarta (BEJ) dengan Bursa Efek Surabaya (BES). Demi efektivitas operasional dan transaksi, Pemerintah memutuskan untuk menggabung Bursa Efek Jakarta (BEJ) sebagai pasar saham dengan Bursa Efek Surabaya (BES) sebagai pasar obligasi dan derivatif. Bursa hasil penggabungan ini mulai beroperasi pada 1 Desember 2007. Daftar perusahaan manufaktur yang tercatat di Bursa Efek Indonesia (BEI) terbagi menjadi tiga jenis sektor yaitu sektor industri dasar dan kimia, sektor aneka industri, dan sektor industri barang konsumsi.

Regresi logistik merupakan suatu metode analisis data yang digunakan untuk mencari hubungan antara variabel respon (y) yang bersifat *biner* atau dikotomis dengan variabel prediktor (x) yang bersifat polikotomis (Hosmer dan Lemeshow, 1989). Dalam penelitian ini, variabel respon yang digunakan adalah *financial distress* dan *non-financial distress*, maka digunakan regresi logistik biner yang variabel respon y terdiri atas 2 kategori. Regresi logistik juga merupakan metode klasifikasi dengan pendekatan parametrik, yang memiliki kelebihan yaitu nilai *odds ratio* yang menunjukkan seberapa besar pengaruh variabel prediktor suatu kategori preferensi pada suatu variabel respon. *Classification and Regression Trees* (CART) adalah salah satu metode atau algoritma dari salah satu teknik eksplorasi data yaitu teknik pohon keputusan. Menurut Breiman (1993), CART merupakan metode statistik dengan pendekatan nonparametrik yang dikembangkan untuk analisis klasifikasi yang mampu mengatasi keterbatasan asumsi. Terdapat beberapa kelebihan

yang dimiliki oleh metode CART yaitu mampu bekerja pada dimensi data yang besar dan struktur data yang kompleks, dapat mengetahui interaksi antar variabel prediktor dan hasil klasifikasi yang diperoleh lebih mudah dipahami serta diinterpretasikan. Oleh karena itu, digunakan pendekatan parametrik dengan metode regresi logistik biner dan pendekatan nonparametrik dengan menggunakan metode CART yang diharapkan dapat membantu menyelesaikan pengklasifikasian *financial distress* perusahaan manufaktur serta faktor-faktor yang berpengaruh terhadap kondisi keuangan.

Selama ini peneliti sebelumnya sudah pernah melakukan klasifikasi *financial distress* dengan berbagai metode yaitu *Support Vector Machine (SVM)* untuk model prediksi *Financial distress* Pada Perusahaan Manufaktur Go Public Di Indonesia oleh Zhahrotun 2013. *Financial distress Prediction Using Particle Swarm Optimization-Support Vector Machine* oleh Herlina tahun 2014, serta Klasifikais *Financial distress* Perusahaan Manufaktur yang Terdaftar di BEI Tahun 2009-2013 Menggunakan Regresi Logistik Biner dan *SVM* oleh Indah tahun 2015.

1.2 Rumusan Masalah

Berdasarkan latar belakang di atas menunjukkan adanya masalah penelitian mengenai faktor-faktor apa sajakah yang dapat mempengaruhi kejadian *financial distress*. Financial distres dapat diakibatkan oleh beberapa penyebab yang bermacam-macam, dapat dipengaruhi oleh pengaruh dari dalam perusahaan sendiri maupun pengaruh dari luar perusahaan. Namun, pada hakikatnya semua perusahaan berusaha agar tidak mengalami *financial distress*. Untuk itu pada penelitian kali ini, akan dilakukan analisis faktor-faktor yang mempengaruhi perusahaan manufaktur mengalami *financial distress* serta pengklasifikasian apakah perusahaan manufaktur yang tercatat di BEI pada tahun 2012 sampai tahun 2014 mengalami *financial distress* atau *non-*

financial distress dari beberapa faktor penduga dengan menggunakan metode regresi logistik biner dan CART.

1.3 Tujuan Penelitian

Berdasarkan perumusan masalah diatas, tujuan dari penelitian yang terbentuk adalah:

1. Mengetahui faktor-faktor yang diduga berpengaruh signifikan terhadap *financial distress*.
2. Mengetahui model klasifikasi *financial distress*, apakah termasuk dalam kondisi *financial distress* atau kondisi *non-financial distress*.

1.4 Manfaat Hasil Penelitian

Manfaat penyusunan tugas akhir ini diharapkan dapat memberikan gambaran tentang kondisi *financial distress* pada perusahaan manufaktur serta dapat memberikan informasi, dan sebagai pertimbangan pada perusahaan dalam mengantisipasi kondisi perusahaan yang mengarah pada kebangkrutan melalui faktor-faktor yang mempengaruhi kondisi *financial distress* dan klasifikasi kondisi *financial distress*.

1.5 Batasan Masalah

Batasan masalah yang terdapat pada penelitian ini yaitu data perusahaan *go public* dari Bursa Efek Indonesia hanya diambil pada sektor industri manufaktur. Variabel prediktor yang digunakan adalah rasio keuangan yang didapat dari laporan keuangan pada tahun 2012 dan 2013 sedangkan variabel respon pada tahun 2013 dan 2014.

(Halaman ini sengaja dikosongkan)

BAB II

TINJAUAN PUSTAKA

2.1 SMOTE

Salah satu metode yang dapat digunakan untuk mengatasi masalah *imbalanced data* yaitu *Synthetic Minority Oversampling Technique* (SMOTE). Dimana SMOTE adalah salah satu metode *oversampling* yang ditemukan oleh Chawla (2003) berupa teknik penambahan jumlah sampel pada kelas minor dengan melakukan replikasi data pada kelas minor secara acak sehingga menghasilkan jumlah data yang sama dengan data pada kelas data mayor. Data yang direplikasi merupakan data yang berasal dari kelas minor. Metode yang digunakan pada algoritma SMOTE adalah *k-nearest neighbors* (ketetanggaan data yang terdekat) yang termasuk dalam kelompok metode statistik nonparametrik. Metode ini bekerja dengan mengelompokkan data terdekat yang dipilih berdasarkan jarak *Euclidean* antara kedua data. Penentuan jumlah replikasi yang dilakukan disesuaikan dengan jumlah anggota pada kelas mayor. Jumlah replikasi harus sesuai dengan jumlah k pada *nearest neighbour*, jika jumlah replikasi sebanyak n maka jumlah k sebanyak $n-1$.

Misalkan terdapat dua struktur data dengan p dimensi yaitu $x^T = [x_1, x_2, \dots, x_p]$ dan $y^T = [y_1, y_2, \dots, y_p]$ maka jarak *Euclidean* $d(x, y)$ yang dihasilkan antara kedua data ditunjukkan pada persamaan sebagai berikut.

$$d(x, y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + \dots + (x_p - y_p)^2} \quad (2.1)$$

Replikasi data dilakukan dengan menggunakan persamaan sebagai berikut.

$$x_{syn} = x_i + (x_{knn} - x_i) \times \tau \quad (2.2)$$

dengan

x_{syn} = data hasil replikasi.

x_i = data yang akan direplikasi

x_{knn} = data yang memiliki jarak terdekat dari data yang akan direplikasi.

τ = bilangan random 0 sampai 1

2.2 Regresi Logistik Biner

Menurut Hosmer dan Lemeshow (1989), regresi logistik merupakan suatu metode analisis data yang digunakan untuk mencari hubungan antara variabel respon (y) yang bersifat *biner* atau dikotomis dengan variabel prediktor (x) yang bersifat polikotomis. *Outcome* dari variabel respon y terdiri dari 2 kategori yaitu “sukses” dan “gagal” yang dinotasikan dengan $y=1$ (sukses) dan $y=0$ (gagal). Dalam keadaan demikian, variabel y mengikuti distribusi *Bernoulli* untuk setiap observasi tunggal. Bentuk umum model regresi logistiknya adalah sebagai berikut.

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}}{1 + e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}} \quad (2.3)$$

dimana: p = banyaknya variabel prediktor

Untuk mempermudah pendugaan parameter regresi maka model regresi logistik pada persamaan (2.3) dapat diuraikan dengan menggunakan transformasi logit dari $\pi(x)$.

$$\begin{aligned} \{\pi(x)\} \left\{ 1 + e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \right\} &= e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \\ \{\pi(x)\} + \{\pi(x) e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)}\} &= e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \\ \pi(x) &= e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} - \pi(x) e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \\ \pi(x) &= \{1 - \pi(x)\} e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \\ \frac{\pi(x)}{1 - \pi(x)} &= e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \\ \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) &= \ln e^{(\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p)} \\ \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) &= \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \end{aligned}$$

Sehingga diperoleh persamaan berikut.

$$g(x) = \ln \left(\frac{\pi(x)}{1 - \pi(x)} \right) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p \quad (2.4)$$

dimana:

$\pi(x)$: peluang kejadian sukses dengan probabilitas $0 \leq \pi(x) \leq 1$

β_i : nilai parameter dengan $i = 1, 2, \dots, p$

2.2.1 Estimasi Parameter

Estimasi parameter dalam regresi logistik dilakukan dengan metode *Maximum Likelihood*. Metode tersebut mengestimasi parameter β dengan cara memaksimumkan fungsi likelihood dan mensyaratkan bahwa data harus mengikuti suatu distribusi tertentu. Apabila, diasumsikan antara amatan yang satu dengan yang lain saling bebas, maka fungsi *Maximum Likelihood Estimation* (MLE) yang diperoleh adalah:

$$l(\beta) = \prod_{i=1}^p \pi(x_i)^{y_i} [1 - \pi(x_i)]^{1-y_i} \quad (2.5)$$

dengan:

$i = 1, 2, \dots, p$

y_i = pengamatan pada variabel respon ke- i

π_i = peluang untuk variabel respon ke- i

Parameter β_i diduga dengan memaksimumkan persamaan diatas (2.5), untuk mempermudah perhitungan maka dilakukan pendekatan logaritma, sehingga fungsi likelihoodnya didapatkan sebagai berikut.

$$\begin{aligned} L(\beta) &= \ln l(\beta) \\ &= \sum_{i=1}^p \{ y_i \ln [\pi(x_i)] + (1 - y_i) \ln [1 - \pi(x_i)] \} \end{aligned} \quad (2.6)$$

Untuk menemukan nilai dugaan β_i dapat diperoleh dengan membuat turunan pertama $L(\beta)$ terhadap $\beta_i = 0$ dimana $i=1, 2, \dots, p$. Persamaan tersebut diketahui sebagai *likelihood equations* yaitu.

$$\sum [y_i - \pi(x_i)] = 0 \quad (2.7)$$

dan

$$\sum x_i [y_i - \pi(x_i)] = 0 \quad (2.8)$$

2.2.2 Pengujian Parameter

Setelah parameter hasil estimasi diperoleh, selanjutnya dilakukan pengujian terhadap parameter-parameter model yang dilakukan secara serentak maupun parsial. Menurut Hosmer dan Lemeshow (1989), pengujian parameter secara serentak dilakukan dengan menggunakan uji nisbah kemungkinan (*Likelihood Ratio Test*). Pengujian ini dilakukan untuk memeriksa keberartian koefisien β secara serentak (multivariat) terhadap variabel respon. Hipotesis yang digunakan diberikan sebagai berikut.

$$H_0: \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1: \text{minimal terdapat satu } \beta_i \neq 0; \quad i = 1, 2, \dots, p$$

$$\text{Statistik uji: } G = -2 \ln \frac{\left(\frac{n_1}{n}\right)^{n_1} \left(\frac{n_0}{n}\right)^{n_0}}{\sum_{i=1}^n \hat{\pi}_i^{y_i} (1 - \hat{\pi}_i)^{(1-y_i)}} \quad (2.9)$$

$$\text{dimana: } n_1 = \sum_{i=1}^n y_i \quad n_0 = \sum_{i=1}^n (1 - y_i)$$

$$n = n_1 + n_0$$

Statistik uji G adalah merupakan *Likelihood Ratio Test* dimana nilai G mengikuti distribusi *Chi-Square* sehingga H_0 ditolak jika $G > \chi^2_{(v, \alpha)}$ dengan v derajat bebas adalah banyaknya parameter dalam model tanpa β_0 .

Pengujian parameter secara parsial menggunakan statistik uji Wald. Hipotesis pengujian parsial adalah sebagai berikut,

$$H_0: \beta_i = 0$$

$$H_1: \beta_i \neq 0 \quad ; \quad i = 1, 2, \dots, p$$

Statistik uji:

$$W_i = \frac{\hat{\beta}_i}{SE(\hat{\beta}_i)} \quad (2.10)$$

Statistik uji W tersebut mengikuti distribusi normal sehinggadiperoleh keputusan H_0 ditolak jika $|W| > Z_{\alpha/2}$ dan jika Statistik uji Wald tersebut mengikuti distribusi *Chi-Squire* dapat diperoleh melalui persamaan berikut,

$$W_i^2 = \frac{\hat{\beta}_i^2}{SE(\hat{\beta}_i)^2} \quad (2.11)$$

Keputusan H_0 ditolak jika $W^2 > \chi^2_{(v,\alpha)}$ dengan v *degrees of freedom* adalah banyaknya variabel prediktor.

2.3 Classification Analysis And Regression Tree (CART)

CART merupakan metodologi statistik nonparametrik yang dikembangkan untuk topik analisis klasifikasi, baik untuk variabel respon kategorik maupun kontinu. CART menghasilkan suatu pohon klasifikasi jika variabel responnya kategorik, dan menghasilkan pohon regresi jika variabel responnya kontinu. Tujuan utama CART adalah mendapatkan suatu kelompok data yang akurat sebagai penciri dari suatu pengklasifikasian. Metode CART merupakan metode yang bisa diterapkan untuk himpunan data yang mempunyai jumlah besar, variabel yang sangat banyak dan dengan skala variabel campuran melalui prosedur pemilahan biner. Langkah-langkah penerapan Algoritma CART adalah sebagai berikut.

1. Pembentukan pohon klasifikasi
- Proses pembentukan pohon klasifikasi terdiri atas 3 tahapan, yaitu
- a. Pemilahan pemilah (*Classifier*)

Pemilahan pemilah tergantung pada jenis pohon atau tergantung pada jenis variabel responnya. Pengukuran tingkat keheterogenan suatu kelas dari suatu simpul tertentu dalam pohon klasifikasi disebut *impurity measure i (t)*.

Ukuran ini akan membantu menemukan fungsi pemilah yang optimum. Fungsi keheterogenan $i(t)$ adalah sebagai berikut.

- i. Indeks Gini: $i(t) = \sum_{i \neq j} P(i|t)P(j|t)$
- ii. Indeks Informasi: $i(t) = -\sum_j P(j|t) \log[P(j|t)]$
- iii. Indeks Twoing: $i(t) = \frac{P_L P_R}{4} [\sum |P(j|t_L) - P(j|t_R)|]^2$
- iv. Indeks Entropi: $i(t) = N_j(t) \log[P(j|t)]$

dimana $P(j|t)$ adalah peluang j pada simpul t .

Goodness of Split $\phi(s, t)$ didefinisikan sebagai penurunan keheterogenan.

$$\phi(s, t) = \Delta i(s, t) = i(t) - P_L i(t_L) - P_R i(t_R) \quad (2.12)$$

Pemilah yang menghasilkan nilai $\Delta i(s, t)$ lebih tinggi, merupakan pemilah yang lebih baik. t_L dan t_R , merupakan partisi dari simpul t menjadi dua himpunan bagian yang saling lepas dimana p_L dan p_R adalah proporsi masing-masing peluang simpul. Karena $t_L \cup t_R = t$, maka nilai $\Delta i(s, t)$ merepresentasikan perubahan dari keheterogenan dalam simpul t yang semata-mata disebabkan oleh pemilah s . Jika simpul yang diperoleh merupakan kelas yang tidak homogen, prosedur yang sama diulangi sampai pohon klasifikasi menjadi suatu konfigurasi tertentu, dan memenuhi.

$$\Delta i(s^*, t_1) = \max_{s \in S} \Delta i(s, t_1) \quad (2.13)$$

- b. Penentuan simpul terminal

Suatu simpul t akan menjadi simpul terminal atau tidak, akan dipilah kembali bila pada simpul t tidak terdapat penurunan keheterogenan dengan adanya batasan minimum n seperti hanya terdapat satu pengamatan pada tiap simpul anak.

c. Penandaan label kelas

Penandaan label kelas pada simpul terminal dilakukan berdasarkan aturan jumlah terbanyak, yaitu

$$P(j_0 | t) = \max_j P(j | t) = \max_j \frac{N_j(t)}{N(t)} \quad (2.14)$$

dengan $P(j_0 | t)$ adalah proporsi kelas j pada simpul t , $N_j(t)$ adalah jumlah pengamatan kelas j pada simpul t dan $N(t)$ adalah jumlah pengamatan pada simpul t . Label kelas simpul terminal t adalah j_0 yang memberi nilai dugaan kesalahan pengklasifikasian simpul t terbesar.

Proses pembentukan pohon klasifikasi berhenti saat terdapat hanya satu pengamatan dalam tiap tiap simpul anak atau adanya batasan minimum n , semua pengamatan dalam tiap simpul anak identik, dan adanya batasan jumlah level/kedalaman pohon maksimum.

2. Pemangkasan pohon klasifikasi

Pemangkasan dilakukan dengan jalan memangkas bagian pohon yang kurang penting sehingga didapatkan pohon optimum. Ukuran pemangkasan yang digunakan untuk memperoleh ukuran pohon yang layak adalah Cost complexity minimum. Sub pohon dari pohon terbesar T_{\max} ($T < T_{\max}$) ukuran *cost complexity* yaitu

$$R\alpha(t) = R(T) + \alpha |\overline{T}| \quad (2.15)$$

dimana

$R(T)$ = Resubstitution Estimate (Proporsi kesalahan pada sub pohon)

α = kompleksitas parameter

$|\overline{T}|$ = ukuran banyaknya simpul terminal pohon T

$R_\alpha(t)$ merupakan kombinasi linear biaya dan kompleksitas pohon yang dibentuk dengan menambahkan cost penalty bagi kompleksitas terhadap biaya kesalahan klasifikasi pohon. *Cost complexity pruning* menentukan suatu pohon bagian $T(\alpha)$ yang meminimumkan pada seluruh pohon bagian. Atau untuk setiap

nilai α , dicari pohon bagian $T(\alpha) < T_{\max}$ yang meminimumkan $R_\alpha(t)$ yaitu

$$R_\alpha(T(\alpha)) = \min_{T < T_{\max}} R_\alpha(T) \quad (2.16)$$

Jika $R(T)$ digunakan sebagai kriteria penentuan pohon optimum maka akan cenderung pohon terbesar adalah T_1 , sebab semakin besar pohon, maka semakin kecil nilai $R(T)$ nya.

3. Penentuan pohon klasifikasi optimum

Pohon klasifikasi yang berukuran besar memberikan nilai penduga pengganti paling kecil, sehingga pohon ini cenderung dipilih untuk menduga nilai respon. Ukuran pohon yang besar akan menyebabkan nilai kompleksitas yang tinggi karena struktur data yang digambarkan cenderung kompleks, sehingga perlu dipilih pohon optimum yang berukuran sederhana tetapi memberikan nilai penduga pengganti cukup kecil.

Sebagai contoh, akan dilakukan ilustrasi pembentukan pohon klasifikasi dengan menggunakan metode CART. Variabel yang digunakan dalam ilustrasi lebih lengkapnya dijelaskan pada Tabel 2.1. Data yang digunakan dalam ilustrasi berjumlah 10 sampel pengamatan yang diambil secara acak, yang ditampilkan pada Tabel 2.2 berikut.

Tabel 2.1 Variabel Ilustrasi

Variabel	Nama Variabel	Kategori	Skala Pengukuran
Y	<i>Financial distress</i> (FD)	0 = tidak terjadi FD 1 = terjadi FD	Nominal
X_1	Rasio Likuiditas	-	Rasio
X_2	Rasio Profitabilitas	-	Rasio

Tabel 2.2 Data Sampel untuk Ilustrasi

Sampel Ke-	Y	X_1	X_2	Sampel Ke-	Y	X_1	X_2
1	0	0.02	2.4	6	0	0.14	2.6
2	0	0.02	2.4	7	1	0.15	2.6
3	0	0.09	2.4	8	0	0.15	2.7
4	1	0.09	2.4	9	0	0.17	2.7
5	0	0.14	2.6	10	1	0.17	2.9

Data sampel sebanyak 10 pengamatan tersebut nantinya akan membentuk suatu simpul yang disebut dengan simpul awal yang kemudian akan dipilah. Sebelum melakukan pemilahan pada simpul awal, terlebih dahulu dihitung nilai keheterogenan simpul awal dengan menggunakan fungsi keheterogenan Indeks Gini. Kemudian dilakukan penentuan pemilah dari semua kemungkinan variabel dan nilai *threshold* untuk memilah simpul awal menjadi dua simpul anak dengan terlebih dahulu menghitung nilai *Goodness of Split* dari masing-masing kemungkinan *threshold* tiap variabel. *Threshold* yang dipilih adalah yang memiliki nilai *Goodness of Split* paling tinggi dimana nilai *threshold* tersebut akan dijadikan sebagai pemilah terbaik untuk simpul awal. Berikut diberikan contoh perhitungan dari proses pemilahan terbaik.

Perhitungan Indeks Gini untuk variabel Y

Kategori 0 = tidak terjadi *financial distress*; 1= terjadi *financial distress*

$$N(t) = 10 \text{ pengamatan}$$

$$N_i(t) = 7 \text{ pengamatan} \rightarrow P(i|t) = 7/10 = 0,7$$

$$N_j(t) = 3 \text{ pengamatan} \rightarrow P(j|t) = 3/10 = 0,3$$

$$i(t) = \sum_{j \neq i} P(j|t)P(k|t)$$

$$i(t) = P(1|t)P(2|t) + P(2|t)P(1|t)$$

$$i(t) = (0,7)(0,3) + (0,3)(0,7) = 0,42$$

Penentuan *threshold* dan perhitungan *Goodness of Split* variabel X_1 (Rasio Likuiditas)

Variabel X_1 merupakan data bertipe kontinu, sesuai dengan aturan dalam pemilahan maka *threshold* dari variabel X_1 adalah rata-rata dari dua nilai amatan sampel berurutan yang berbeda dari variabel X_1 . Berikut ditunjukkan perhitungan rata-rata untuk menentukan *threshold*.

Tabel 2.3 Rata-rata Dua Amatan Sampel Berurutan Variabel X_1

No.	Y	X_1	
1	0	0.02	0,055
2	0	0.02	
3	0	0.09	
4	1	0.09	0,145
5	0	0.14	
6	0	0.14	0,115
7	1	0.15	
8	0	0.15	0,160
9	0	0.17	
10	1	0.17	

Rata-rata dari dua amatan sampel berurutan tersebut, kemudian akan menjadi nilai *threshold* dan selanjutnya akan dilakukan perhitungan nilai *Goodness of Split* untuk masing-masing *threshold* sebagai berikut.

Tabel 2.4 Perhitungan *Goodness of Split* untuk *Threshold* 0,055

Simpul Kiri		Simpul Kanan	
$N_L = 2$	$P_L=0,2$	$N_R = 8$	$P_R = 0,8$
$N(0 t_L) = 2$	$P(0 t_L) = 1$	$N(0 t_R) = 5$	$P(0 t_R) = 0,63$
$N(1 t_L) = 0$	$P(1 t_L) = 0$	$N(1 t_R) = 3$	$P(1 t_R) = 0,37$
$i(t_L) = 0$		$i(t_R) = 0,23$	
$Goodness\ of\ Split = \phi(s,t) = 0,2325$			

Tabel 2.5 Perhitungan *Goodness of Split* untuk *Threshold* 0,115

Simpul Kiri		Simpul Kanan	
$N_L = 4$	$P_L = 0,4$	$N_R = 6$	$P_R = 0,6$
$N(0 t_L) = 3$	$P(0 t_L) = 0,75$	$N(0 t_R) = 4$	$P(0 t_R) = 0,67$
$N(1 t_L) = 1$	$P(1 t_L) = 0,25$	$N(1 t_R) = 2$	$P(1 t_R) = 0,33$
$i(t_L) = 0,18$		$i(t_R) = 0,22$	
$Goodness\ of\ Split = \phi(s,t) = 0,2117$			

Tabel 2.6 Perhitungan *Goodness of Split* untuk *Threshold* 0,145

Simpul Kiri		Simpul Kanan	
$N_L = 6$	$P_L = 0,6$	$N_R = 4$	$P_R = 0,4$
$N(0 t_L) = 5$	$P(0 t_L) = 0,83$	$N(0 t_R) = 2$	$P(0 t_R) = 0,5$
$N(1 t_L) = 1$	$P(1 t_L) = 0,17$	$N(1 t_R) = 2$	$P(1 t_R) = 0,5$
$i(t_L) = 0,14$		$i(t_R) = 0,25$	
$Goodness\ of\ Split = \phi(s,t) = 0,2117$			

Tabel 2.7 Perhitungan *Goodness of Split* untuk *Threshold* 0,160

Simpul Kiri		Simpul Kanan	
$N_L = 8$	$P_L = 0,8$	$N_R = 2$	$P_R = 0,2$
$N(0 t_L) = 6$	$P(0 t_L) = 0,75$	$N(0 t_R) = 1$	$P(0 t_R) = 0,5$
$N(1 t_L) = 2$	$P(1 t_L) = 0,25$	$N(1 t_R) = 1$	$P(1 t_R) = 0,5$
$i(t_L) = 0,18$		$i(t_R) = 0,25$	
<i>Goodness of Split</i> = $\phi(s,t) = 0,2222$			

Penentuan *threshold* dan perhitungan *Goodness of Split* variabel X_2 (Rasio Profitabilitas)

Variabel X_2 merupakan data bertipe kontinu, sesuai dengan aturan dalam pemilahan maka *threshold* dari variabel X_2 adalah rata-rata dari dua nilai amatan sampel berurutan yang berbeda dari variabel X_2 . Berikut ditunjukkan perhitungan rata-rata untuk menentukan *threshold*.

Tabel 2.8 Rata-rata Dua Amatan Sampel Berurutan Variabel X_2

No.	Y	X_2	
1	0	2.4	2,49
2	0	2.4	
3	0	2.4	
4	1	2.4	
5	0	2.6	2,63
6	0	2.6	
7	1	2.6	2,74
8	0	2.7	
9	0	2.7	
10	1	2.9	

Rata-rata dari dua amatan sampel berurutan tersebut, kemudian akan menjadi nilai *threshold* dan selanjutnya akan dilakukan perhitungan nilai *Goodness of Split* untuk masing-masing *threshold* sebagai berikut.

Tabel 2.9 Perhitungan *Goodness of Split* untuk *Threshold* 2,49

Simpul Kiri		Simpul Kanan	
$N_L = 4$	$P_L=0,4$	$N_R = 6$	$P_R = 0,6$
$N(0 t_L) = 3$	$P(0 t_L) = 0,75$	$N(0 t_R) = 4$	$P(0 t_R) = 0,67$
$N(1 t_L) = 1$	$P(1 t_L) = 0,25$	$N(1 t_R) = 2$	$P(1 t_R) = 0,33$
$i(t_L) = 0,19$		$i(t_R) = 0,22$	
<i>Goodness of Split</i> = $\phi(s,t) = 0,2116$			

Tabel 2.10 Perhitungan *Goodness of Split* untuk *Threshold* 2,63

Simpul Kiri		Simpul Kanan	
$N_L = 7$	$P_L=0,7$	$N_R = 3$	$P_R = 0,3$
$N(0 t_L) = 5$	$P(0 t_L) = 0,71$	$N(0 t_R) = 2$	$P(0 t_R) = 0,67$
$N(1 t_L) = 2$	$P(1 t_L) = 0,29$	$N(1 t_R) = 1$	$P(1 t_R) = 0,33$
$i(t_L) = 0,20$		$i(t_R) = 0,22$	
$Goodness\ of\ Split = \phi(s,t) = 0,2105$			

Tabel 2.11 Perhitungan *Goodness of Split* untuk *Threshold* 2,74

Simpul Kiri		Simpul Kanan	
$N_L = 9$	$P_L=0,9$	$N_R = 1$	$P_R = 0,1$
$N(0 t_L) = 7$	$P(0 t_L) = 0,78$	$N(0 t_R) = 0$	$P(0 t_R) = 0$
$N(1 t_L) = 2$	$P(1 t_L) = 0,22$	$N(1 t_R) = 1$	$P(1 t_R) = 1$
$i(t_L) = 0,17$		$i(t_R) = 0$	
$Goodness\ of\ Split = \phi(s,t) = 0,2644$			

Penentuan pemilah terbaik

Pemilah terbaik merupakan *threshold* variabel yang memiliki nilai *Goodness of Split* paling besar. Perbandingan nilai *Goodness of Split* dari semua kemungkinan pemilah ditunjukkan pada Tabel 2.12.

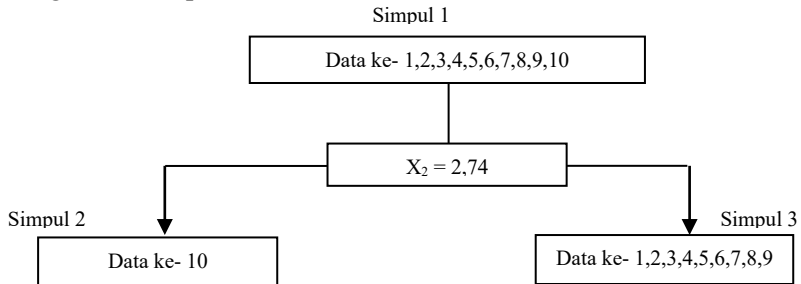
Tabel 2.12 Perbandingan *Goodness of Split* dari Setiap Kemungkinan Pemilah Satu

Variabel	<i>Threshold</i>	<i>Goodness of Split</i>
X_1	0,055	0,2325
	0,115	0,2117
	0,145	0,2117
	0,160	0,2222
X_2	2,49	0,2116
	2,63	0,2105
	2,74	0,2644*

**Goodness of Split* terbesar

Perbandingan nilai *Goodness of Split* dari semua kemungkinan pemilah untuk variabel X_1 dan X_2 seperti dicantumkan pada Tabel 2.12 menunjukkan bahwa nilai *Goodness of Split* yang paling besar adalah pada variabel X_2 dengan *threshold* sebesar 2,74. Maka, variabel rasio profitabilitas (X_2) akan menjadi pemilah utama pada ilustrasi ini. Kemudian, data sampel yang dimiliki dibagi menjadi dua bagian berdasarkan

pemilah utama. Data sampel yang memiliki nilai X_2 kurang dari *threshold* 2,74 masuk dalam simpul kiri sedangkan data dengan nilai X_2 lebih dari 2,74 akan masuk dalam simpul kanan. Hasil pohon yang terbentuk dari proses pemilahan tersebut digambarkan pada Gambar 2.1.



Gambar 2.1 Pohon Klasifikasi Hasil Pemilahan Simpul Awal

2.4 *Financial distress*

Financial distress adalah suatu kondisi dimana perusahaan menghadapi masalah kesulitan keuangan. Menurut Platt dan Platt (2002) dalam Ellen dan Juniarti (2013) *financial distress* didefinisikan sebagai tahap penurunan kondisi keuangan yang terjadi sebelum terjadinya kebangkrutan ataupun likuidasi. Kondisi *financial distress* tergambar dari ketidak mampuan atau tidak tersedianya dana untuk membayar kewajiban yang telah jatuh tempo. Kesulitan keuangan dapat diartikan dalam beberapa kategori yaitu sebagai berikut :

1. *Economic Failure*, yaitu kegagalan ekonomi yang berarti bahwa pendapatan perusahaan tidak dapat menutup biaya total. Ini berarti tingkat laba lebih kecil dari biaya modal.
2. *Bussines Failure*, didefenisikan sebagai usaha yang menghentikan operasinya dengan akibat kerugian bagi kreditur, dan kemudian dikatakan dengan akibat kerugian bagi kreditur, dan kemudian dikatakan gagal meskipun tidak melalui kebangkrutan secara normal.

3. *Technical insolvency*, sebuah perusahaan dapat dinilai mengalami kesulitan keuangan apabila tidak memenuhi kewajibannya yang jatuh tempo. *Technical insolvency* ini menunjukkan kekurangan likuiditas yang sifatnya sementara dimana pada suatu waktu perusahaan dapat mengumpulkan uang untuk memenuhi kewajibannya dan tetap beroperasi.
4. *Insolvency in bankruptcy*, sebuah perusahaan dapat dikatakan mengalami kesulitan keuangan bilamana nilai buku dari total kewajiban melebihi nilai pasar dari asset perusahaan.
5. *Legal Bankruptcy*, sebuah perusahaan dikatakan bangkrut secara hukum, kecuali diajukan tuntutan secara resmi dengan undang-undang federal.

Suatu perusahaan dapat dikategorikan sedang mengalami *financial distress* jika perusahaan tersebut memiliki kinerja yang menunjukkan laba operasinya negatif, laba bersih negatif, dan nilai buku ekuitas negatif (Brahmana, 2007). Ciccone (2001) dalam (Ellen dan Juniarti, 2013) memaparkan bahwa ada dua *proxy* yang digunakan untuk mengukur kesulitan keuangan yaitu kerugian didefinisikan ketika laba sebenarnya pada tahun n kurang dari 0 dan penurunan laba didefinisikan sebagai pendapatan tahun n lebih kecil dari pada pendapatan tahun $n-1$. Kriteria-kriteria yang telah disebutkan diatas adalah kriteria *financial distress* yang akan digunakan dalam penelitian ini yaitu *net income* negatif dan *retained earning* negatif. Peraturan yang mengatur hubungan antara pemegang saham, pengurus (pengelola) perusahaan, pihak kreditur, pemerintah, karyawan serta para pemegang kepentingan intern dan ekstern lainnya yang berkaitan dengan hak-hak dan kewajiban mereka, atau dengan kata lain suatu sistem yang mengatur dan mengendalikan perusahaan.

2.5 Rasio Keuangan Model Altman

Rasio keuangan yang digunakan dalam penelitian ini adalah variabel rasio keuangan yang digunakan dalam penelitian Altman (1968), yaitu:

1. Rasio Modal Kerja terhadap Total Aktiva

Rasio ini disebut juga rasio *Working Capital to Total Assets* atau rasio WC/TA yang merupakan rasio untuk mengukur aktiva lancar bersih perusahaan relatif terhadap total kapitalisnya. Modal kerja didefinisikan sebagai selisih antara aktiva lancar dan kewajiban lancar. Likuiditas dan karakteristik ukuran dipertimbangkan secara jelas. Pada semua kasus semua aktiva berwujud, tidak termasuk aktiva tak berwujud digunakan dalam perhitungan (Altman dan Hotchkiss, 2006 dalam Hatoyo 2013). Rasio ini dihitung dengan cara:

$$\text{Rasio WC/TA} = \text{Working Capital} / \text{Total Assets}$$

2. Rasio Laba Ditahan terhadap Total Aktiva

Rasio laba ditahan terhadap total aktiva atau *Retained Earnings To Total Assets* (RE/TA) mengukur tingkat *leverage* sebuah perusahaan. Perusahaan yang memiliki rasio RE/TA tinggi membiayai aktivasnya melalui retensi laba dan tidak banyak menggunakan hutang. Rasio ini menunjukkan perbedaan antara penggunaan dana sendiri/laba ditahan (modal beresiko rendah) atau OPM (*Other People's Money*)/hutang (modal beresiko tinggi). Rasio ini dihitung dengan cara:

$$\text{Rasio RE/TA} = \text{Retained Earnings} / \text{Total Assets}$$

3. Rasio EBIT Terhadap Total Aktiva

Earning power of total investment rate to total assets atau rasio EBIT/TA mengukur produktivitas aktiva perusahaan tanpa pengaruh dari tingkat pajak ataupun tingkat bunga. Karena eksistensi perusahaan dinilai berdasarkan kemampuan aktiva menghasilkan laba. Rasio ini dihitung dengan cara:

$$\text{Rasio EBIT/TA} = \text{EIBIT} / \text{Total Assets}$$

4. Rasio Nilai Pasar Ekuitas terhadap Nilai Buku

Nilai pasar ekuitas yang dimaksud adalah ngabungan dari nilai pasar saham biasa dan nilai pasar saham preferen. Rasio ini menunjukkan seberapa jauh nilai pasar ekuias perusahaan

akan menurun sebelum kewajiban melebihi aktiva sehingga perusahaan menjadi insolven. Rasio ini digunakan dengan cara:

$$\text{Rasio MVE/BVD} = \frac{\text{Market Value of Equity}}{\text{Book Value of Total Debt}}$$

5. Rasio Penjualan terhadap Total Aktiva

Rasio ini merupakan rasio perputaran modal atau rasio standar untuk mengilustrasikan kemampuan menciptakan penjualan dari aktiva yang dimiliki perusahaan. Rasio ini dihitung dengan cara:

$$\text{Rasio A/TA} = \frac{\text{Sales}}{\text{Total Assets}}$$

6. ROI (*Return On Investment*)

Rasio ini mengukur kemampuan perusahaan dengan jumlah aktiva yang digunakan dalam operasi perusahaan untuk menghasilkan keuntungan. Rasio ini menunjukkan produktivitas dari seluruh dana perusahaan baik modal pinjaman maupun modal sendiri. Semakin kecil/rendah rasio ini semakin tidak baik, demikian juga sebaliknya. Rasio ini dihitung dengan cara:

$$\text{ROI} = \frac{\text{Total penjualan-Investasi}}{\text{Investasi}}$$

7. ROE (*Return On Equity*)

ROE adalah rasio yang menunjukkan kemampuan perusahaan dalam menghasilkan laba bersih dengan menggunakan modal sendiri dan menghasilkan laba bersih yang tersedia bagi pemilik atau investor. ROE sangat bergantung pada besar kecilnya perusahaan, misalnya untuk perusahaan kecil tentu memiliki modal yang relative kecil, sehingga ROE yang dihasilkanpun kecil, begitu pula sebaliknya untuk perusahaan besar. Rasio ini dihitung dengan cara:

$$\text{ROE} = \frac{\text{Profit Before Taxes Income}}{\text{Total Equity}}$$

BAB III

METODE PENELITIAN

3.1 Sumber Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang berasal dari laporan tahunan (*annual report*) perusahaan tahun 2012-2014. populasi dalam penelitian ini adalah perusahaan *go public* yang tercatat di Bursa Efek Indonesia (BEI) dan sampel penelitian adalah perusahaan manufaktur *go public* yang tercatat di BEI. Data diperoleh dari BEI, Saham OK dan ICMD (*Indonesian Capital Market Directory*). Kriteria pengambilan sampel perusahaan yang termasuk *financial distress* dan *non-financial distress* adalah sebagai berikut.

1. Perusahaan *financial distress*

Adapun kriteria pemilihan sampel perusahaan *financial distress* adalah sebagai berikut.

- a. Perusahaan manufaktur terdaftar dalam Bursa Efek Indonesia pada tahun 2012-2014.
- b. Perusahaan manufaktur pernah mengalami laba bersih negatif pada tahun sebelumnya.
- c. Perusahaan manufaktur memiliki laporan komite audit yang lengkap yaitu perusahaan yang memiliki laporan pertemuan komite audit dan komposisi komite audit.

2. Perusahaan *non financial distress*

Pada penelitian ini, sampel perusahaan *non financial distress* yang diambil berjumlah sama dengan sampel perusahaan *financial distress*. Adapun kriteria pemilihan sampel perusahaan *non-financial distress* adalah sebagai berikut.

- a. Perusahaan manufaktur tidak pernah mengalami laba bersih negatif pada tahun sebelumnya.
- b. Perusahaan manufaktur menerbitkan laporan keuangan pada tahun 2012-2014.

- c. Perusahaan manufaktur memiliki laporan komite audit yang lengkap yaitu perusahaan yang memiliki laporan pertemuan komite audit dan komposisi komite audit.

3.2 Variabel Penelitian

Variabel yang digunakan di dalam penelitian ini terdiri dari variabel prediktor atau variabel independen (x) dan variabel respon atau variabel dependen (y) sebagai berikut.

Tabel 3.1 Variabel Penelitian

Variabel	Nama Variabel	Pengukuran	Skala
x ₁	Rasio Likuiditas	<i>Working Capital</i> dibagi <i>Total Assets</i>	Rasio
x ₂	Rasio Profitabilitas	<i>Retained Earning</i> dibagi <i>Total Assets</i>	Rasio
x ₃	Rasio Rentabilitas	<i>Earning Before Interest and Taxes</i> dibagi <i>Total Assets</i>	Rasio
x ₄	Rasio Solvabilitas	<i>Market Value Equity</i> dibagi <i>Value Of Total Debt</i>	Rasio
x ₅	Rasio Aktivitas	<i>Sales</i> dibagi <i>Total Assets</i>	Rasio
x ₆	ROI	(Total penjualan-Investasi) dibagi Investasi	Rasio
x ₇	ROE	<i>Profit Before Taxes Income</i> dibagi <i>Total Equity</i>	Rasio
Y	<i>Financial distress</i>	1 jika <i>financial distress</i> , 0 jika <i>non -financial distress</i>	Nominal

3.3 Langkah Analisis

Langkah-langkah analisis yang dilakukan dalam penelitian ini adalah sebagai berikut.

1. Melakukan pengumpulan data sekunder, yaitu data laporan keuangan perusahaan manufaktur *go public* yang terdaftar di BEI tahun 2012-2014.
2. Melakukan klasifikasi awal terhadap data untuk kriteria *financial distress* dan *non financial distress*.
3. Membagi data menjadi data *training* dan data *testing*.

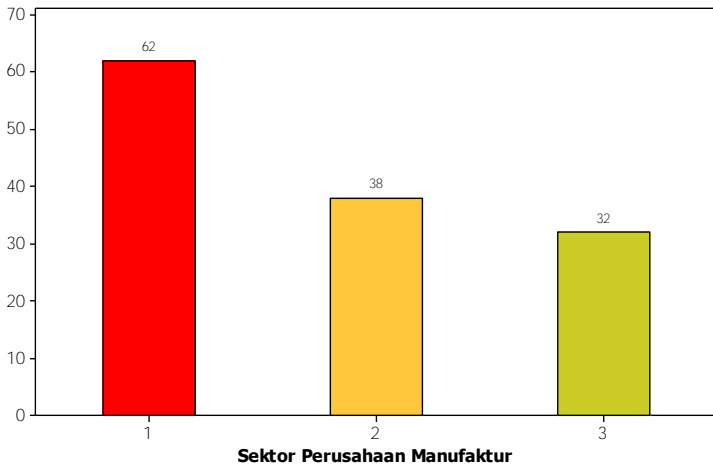
4. Melakukan analisis statistika deskriptif untuk mengetahui karakteristik data pada masing-masing kelas, *financial distress* dan *non-financial distress*.
5. Melakukan analisis SMOTE untuk *imbalanced data*. Tahapannya adalah sebagai berikut:
 - a. Menentukan data minor dan mayor.
 - b. Mereplikasi setiap data minor dengan mencari nilai *k-nearest neighbour*.
 - c. Menghitung sintetik data.
6. Melakukan analisis regresi logistik biner. Tahapannya adalah sebagai berikut:
 - a. Melakukan uji asumsi independensi.
 - b. Estimasi Parameter
 - c. Melakukan pengujian parameter secara serentak menggunakan statistik uji-G.
 - d. Melakukan pengujian parameter secara parsial dengan statistik uji-Wald.
 - e. Memodelkan kondisi keuangan perusahaan berdasarkan variabel prediktor yang berpengaruh.
 - f. Melakukan uji ketepatan klasifikasi.
 - g. Interpretasi hasil
7. Melakukan analisis CART. Tahapannya adalah sebagai berikut.
 - a. Membentuk pohon klasifikasi dengan menentukan pemilihan pemilah, penentuan simpul terminal, dan penandaan label kelas.
 - b. Melakukan pemangkasan pohon klasifikasi.
 - c. Menentukan pohon klasifikasi optimum.
8. Membandingkan hasil dari analisis regresi logistik biner dan metode CART dengan melihat peluang kesalahan klasifikasi.
9. Melakukan kesimpulan dan saran.

(Halaman ini sengaja dikosongkan)

BAB IV HASIL DAN PEMBAHASAN

4.1 Deskriptif Data Perusahaan Manufaktur

Analisis deskriptif bertujuan untuk mengetahui karakteristik data yang digunakan. Variabel yang digunakan dalam penelitian ini meliputi variabel prediktor rasio likuiditas (X_1) sampai ROE (X_7), serta variabel respon *financial distress* (y) dengan kategori 0 dan 1. Berikut merupakan deskriptif data untuk perusahaan manufaktur.



Gambar 4.1 Jumlah Perusahaan Manufaktur per Sektor Manufaktur

Berdasarkan Gambar 4.1 diatas dapat diketahui bahwa sektor industri manufaktur terdiri dari 3 sektor yaitu industri dasar dan kimia, sektor aneka industri, dan sektor industri barang konsumsi. Dimana sebanyak 62 perusahaan merupakan perusahaan manufaktur pada sektor industri dasar dan kimia, 38

perusahaan merupakan sektor aneka industri dan 32 perusahaan lainnya sektor industri barang konsumsi.

Tabel 4.1 Persentase Perusahaan Manufaktur Mengalami *Financial Distress* dan *Non-Financial Distress* per Sektor Manufaktur Tahun 2013-2014

	tidak terjadi <i>financial distress</i>	terjadi <i>financial distress</i>	total
sektor 1	37%	10%	47%
sektor 2	23%	5%	28%
sektor 3	22%	3%	25%
total	83%	17%	100%

Berdasarkan Tabel 4.1 dapat diperoleh informasi bahwa pada sektor 1 yaitu sektor industri dasar dan kimia terdapat 10% perusahaan manufaktur yang mengalami masalah *financial distress*. Kemudian untuk sektor aneka industri terdapat 5% perusahaan manufaktur yang mengalami masalah *financial distress* dan sektor industri barang konsumsi sebesar 3% perusahaan manufaktur yang mengalami masalah *financial distress*. Sehingga, dari 132 perusahaan manufaktur terdapat 17% perusahaan industri manufaktur yang mengalami masalah *financial distress*.

Tabel 4.2 Persentase Perusahaan Manufaktur Mengalami *Financial Distress* dan *Non-Financial Distress* per Sektor Manufaktur Tahun 2012-2013

	tidak terjadi <i>financial distress</i>	terjadi <i>financial distress</i>	total
sektor 1	36%	11%	47%
sektor 2	23%	5%	29%
sektor 3	23%	0,7%	24%
total	83%	16,7%	100%

Berdasarkan Tabel 4.2 dapat diperoleh informasi bahwa pada sektor 1 yaitu sektor industri dasar dan kimia terdapat 11% perusahaan manufaktur yang mengalami masalah *financial distress*. Kemudian untuk sektor aneka industri terdapat 5%

perusahaan manufaktur yang mengalami masalah *financial distress* dan sektor industri barang konsumsi sebesar 0,7% perusahaan manufaktur yang mengalami masalah *financial distress*. Sehingga, dari 132 perusahaan manufaktur terdapat 16,7% perusahaan industri manufaktur yang mengalami masalah *financial distress*.

Tabel 4.3 Statistika Deskriptif Variabel Prediktor per Masing-masing Kelas *Financial Distress* dan *Non-Financial Distress* Tahun 2013-2014

Variable	Kelas	Mean	StDev	Minimum	Maximum
X ₁	0	0,525	0,320	-1,863	0,999
	1	0,118	0,877	-3,302	0,959
X ₂	0	0,288	0,330	0,0019	2,277
	1	0,930	1,663	0,003	7,901
X ₃	0	0,089	0,094	0,0002	0,483
	1	0,064	0,081	0,005	0,297
X ₄	0	1,311	1,773	0,020	13,40
	1	2,197	2,677	0,040	10,90
X ₅	0	1,136	0,931	0,004	7,763
	1	0,991	0,780	0,027	3,473
X ₆	0	8,510	8,237	0,110	47,24
	1	5,66	4,98	0,22	19,07
X ₇	0	13,632	10,359	0,010	57,100
	1	10,01	9,15	0,27	29,20

Ket: rasio likuiditas (X₁), rasio profitabilitas (X₂), rasio rentabilitas (X₃)
Rasio solvabilitas(X₄), rasio aktivitas (X₅), ROI (X₆), ROE (X₇),
Financial Distress (1), *Non-Financial Distress* (0)

Berdasarkan Tabel 4.3 dapat diperoleh informasi bahwa rasio likuiditas (X₁) memiliki rata-rata sebesar 0,525 untuk kategori tidak terjadi *financial distress* (0). Artinya rasio likuiditas sebesar 50% berada diatas 0,525 dan 50% berada dibawah 0,525 dengan batas minimum -1,863 dan batas maksimum 0,999. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 0,118 yang artinya, rasio likuiditas sebesar 50% berada diatas 0,118 dan 50% berada dibawah 0,118 dengan batas minimum -3,302 dan batas

maksimum 0,959. Rasio profitabilitas (X_2) memiliki rata-rata sebesar 0,288 untuk kategori tidak terjadi *financial distress* (0). Artinya rasio profitabilitas sebesar 50% berada diatas 0,288 dan 50% berada dibawah 0,288 dengan batas minimum 0,0019 dan batas maksimum 2,277. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 0,930 yang artinya, rasio profitabilitas sebesar 50% berada diatas 0,930 dan 50% berada dibawah 0,930 dengan batas minimum 0,003 dan batas maksimum 7,901.

Rasio rentabilitas (X_3) memiliki rata-rata sebesar 0,089 untuk kategori tidak terjadi *financial distress* (0). Artinya rasio rentabilitas sebesar 50% berada diatas 0,089 dan 50% berada dibawah 0,089 dengan batas minimum 0,0002 dan batas maksimum 0,483. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 0,064 yang artinya, rasio rentabilitas sebesar 50% berada diatas 0,064 dan 50% berada dibawah 0,064 dengan batas minimum 0,005 dan batas maksimum 0,297. Rasio solvabilitas (X_4) memiliki rata-rata sebesar 1,311 untuk kategori tidak terjadi *financial distress* (0). Artinya rasio solvabilitas sebesar 50% berada diatas 1,311 dan 50% berada dibawah 1,311 dengan batas minimum 0,020 dan batas maksimum 13,40. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 2,197 yang artinya, rasio solvabilitas sebesar 50% berada diatas 2,197 dan 50% berada dibawah 2,197 dengan batas minimum 0,040 dan batas maksimum 10,90.

Rasio aktivitas (X_5) memiliki rata-rata sebesar 1,136 untuk kategori tidak terjadi *financial distress* (0). Artinya rasio aktivitas sebesar 50% berada diatas 1,136 dan 50% berada dibawah 1,136 dengan batas minimum 0,004 dan batas maksimum 7,763. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 0,991 yang artinya, rasio aktivitas sebesar 50% berada diatas 0,991 dan 50% berada dibawah 0,991 dengan batas minimum 0,027 dan batas maksimum 3,473. ROI (X_6) memiliki rata-rata sebesar 8,510

untuk kategori tidak terjadi *financial distress* (0). Artinya ROI sebesar 50% berada diatas 8,510 dan 50% berada dibawah 8,510 dengan batas minimum 0,110 dan batas maksimum 47,24. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 5,66 yang artinya, ROI sebesar 50% berada diatas 5,66 dan 50% berada dibawah 5,66 dengan batas minimum 0,220 dan batas maksimum 19,07. ROE (X_7) memiliki rata-rata sebesar 13,632 untuk kategori tidak terjadi *financial distress* (0). Artinya ROE sebesar 50% berada diatas 13,632 dan 50% berada dibawah 13,632 dengan batas minimum 0,010 dan batas maksimum 57,10. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 10,01 yang artinya, ROI sebesar 50% berada diatas 10,01 dan 50% berada dibawah 10,01 dengan batas minimum 0,27 dan batas maksimum 29,20.

Tabel 4.4 Statistika Deskriptif Variabel Prediktor per Masing-masing Kelas *Financial Distress* dan *Non-Financial Distress* Tahun 2012-2013

Variable	Kelas	Mean	StDev	Minimum	Maximum
X_1	0	0,528	0,308	-1,876	0,992
	1	0,175	0,687	-1,979	0,977
X_2	0	0,339	0,364	0,003	2,581
	1	0,721	1,125	0,002	5,103
X_3	0	0,113	0,10127	0,0014	0,527
	1	0,074	0,0920	0,001	0,458
X_4	0	1,247	1,392	0,090	9,040
	1	2,019	1,551	0,020	6,040
X_5	0	1,208	0,714	0,089	5,659
	1	0,767	0,505	0,002	1,791
X_6	0	8,781	6,964	0,100	32,460
	1	9,24	19,75	0,23	96,81
X_7	0	14,228	9,691	0,110	47,540
	1	14,82	20,43	0,49	99,01

Ket: rasio likuiditas (X_1), rasio profitabilitas (X_2), rasio rentabilitas (X_3)
 Rasio solvabilitas(X_4),rasio aktivitas (X_5),ROI (X_6), ROE (X_7),
Financial Distress (1), *Non-Financial Distress* (0)

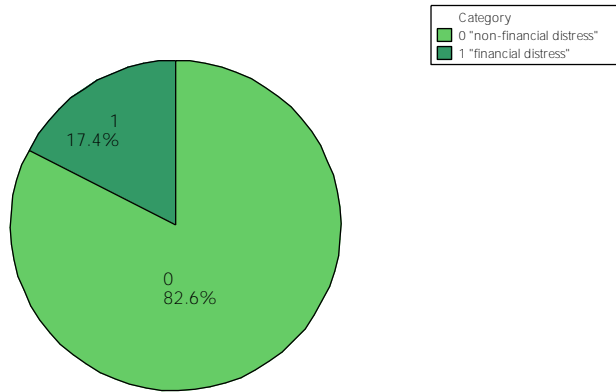
Berdasarkan Tabel 4.4 dapat diperoleh informasi bahwa rasio likuiditas (X_1) memiliki rata-rata sebesar 0,528 untuk kategori tidak terjadi *financial distress* (0). Artinya rasio likuiditas sebesar 50% berada diatas 0,528 dan 50% berada dibawah 0,528 dengan batas minimum -1,876 dan batas maksimum 0,992. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 0,175 yang artinya, rasio likuiditas sebesar 50% berada diatas 0,175 dan 50% berada dibawah 0,175 dengan batas minimum -1,979 dan batas maksimum 0,977. Rasio profitabilitas (X_2) memiliki rata-rata sebesar 0,339 untuk kategori tidak terjadi *financial distress* (0). Artinya rasio profitabilitas sebesar 50% berada diatas 0,339 dan 50% berada dibawah 0,339 dengan batas minimum 0,003 dan batas maksimum 2,581. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 0,721 yang artinya, rasio profitabilitas sebesar 50% berada diatas 0,721 dan 50% berada dibawah 0,721 dengan batas minimum 0,002 dan batas maksimum 5,103.

Rasio rentabilitas (X_3) memiliki rata-rata sebesar 0,113 untuk kategori tidak terjadi *financial distress* (0). Artinya rasio rentabilitas sebesar 50% berada diatas 0,113 dan 50% berada dibawah 0,113 dengan batas minimum 0,0014 dan batas maksimum 0,527. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 0,074 yang artinya, rasio rentabilitas sebesar 50% berada diatas 0,074 dan 50% berada dibawah 0,074 dengan batas minimum 0,001 dan batas maksimum 0,458. Rasio solvabilitas (X_4) memiliki rata-rata sebesar 1,247 untuk kategori tidak terjadi *financial distress* (0). Artinya rasio solvabilitas sebesar 50% berada diatas 1,247 dan 50% berada dibawah 1,247 dengan batas minimum 0,090 dan batas maksimum 9,040. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 2,019 yang artinya, rasio solvabilitas sebesar 50% berada diatas 2,019 dan 50% berada dibawah 2,019 dengan batas minimum 0,020 dan batas maksimum 6,040.

Rasio aktivitas (X_5) memiliki rata-rata sebesar 1,028 untuk kategori tidak terjadi *financial distress* (0). Artinya rasio aktivitas sebesar 50% berada diatas 1,028 dan 50% berada dibawah 1,028 dengan batas minimum 0,089 dan batas maksimum 5,659. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 0,767 yang artinya, rasio aktivitas sebesar 50% berada diatas 0,767 dan 50% berada dibawah 0,767 dengan batas minimum 0,002 dan batas maksimum 1,791. ROI (X_6) memiliki rata-rata sebesar 8,781 untuk kategori tidak terjadi *financial distress* (0). Artinya ROI sebesar 50% berada diatas 8,781 dan 50% berada dibawah 8,781 dengan batas minimum 0,100 dan batas maksimum 32,46. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 9,240 yang artinya, ROI sebesar 50% berada diatas 9,240 dan 50% berada dibawah 9,240 dengan batas minimum 0,230 dan batas maksimum 96,81. ROE (X_7) memiliki rata-rata sebesar 14,23 untuk kategori tidak terjadi *financial distress* (0). Artinya ROE sebesar 50% berada diatas 14,23 dan 50% berada dibawah 14,23 dengan batas minimum 0,110 dan batas maksimum 47,54. Kemudian untuk kategori terjadi *financial distress* (1) memiliki rata-rata sebesar 14,82 yang artinya, ROI sebesar 50% berada diatas 14,82 dan 50% berada dibawah 14,82 dengan batas minimum 0,490 dan batas maksimum 99,01.

4.2 *Preprocessing Imbalanced Data*

Metode yang digunakan untuk *preprocessing imbalanced data* adalah *Synthetic Minority Oversampling Technique* (SMOTE) yang merupakan teknik pengambilan sampel untuk meningkatkan jumlah data pada kelas minoritas dengan cara mereplikasi jumlah data pada kelas minoritas secara acak sehingga jumlahnya sama dengan data pada kelas mayoritas.



Gambar 4.2 Persentase Masing-masing Kelas *Financial Distress* dan *Non-Financial Distress* untuk Tahun 2013-2014

Berdasarkan Gambar 4.2 maka kelas minor merupakan kategori 1 yaitu perusahaan yang mengalami *financial distress* dengan presentase sebesar 17.4%. Tahapan yang dilakukan pada algoritma SMOTE adalah Setiap data pada kelas minor akan direplikasi dengan mencari tetangga terdekat (x_{knn}) menggunakan jarak *euclidean*. Dari perhitungan pada Lampiran 1.3 maka dapat diambil dua jarak *euclidean* yang terdekat pada kelas 1 ($Y=1$ “terjadi *financial distress*”) yaitu $\sqrt{4}$ dan $\sqrt{5}$ maka kelas 1 akan direplikasi empat kali. Jumlah data kelas 1 yang semula berjumlah 23 maka setelah direplikasi sebanyak 4 kali akan menjadi 115 data. Hasil distribusi data simulasi SMOTE, dapat dilihat pada Tabel 4.5 dan data simulasi setelah SMOTE, dapat dilihat pada Tabel 4.6.

Tabel 4.5 Distribusi Data Simulasi Sebelum dan Setelah SMOTE Tahun 2013-2014

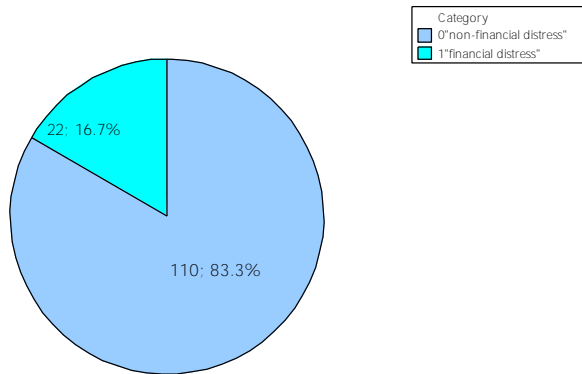
Kelas mayor	Kelas minor	Replikasi	Kelas mayor	Kelas minor baru
109(83%)	23 (17%)	4	109(49%)	115(51%)

Tabel 4.6 Data Simulasi Setelah Menggunakan SMOTE
Tahun 2013-2014

Data ke-	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	Y
1	0,50	0,03	0,06	0,70	0,61	6,76	11,47	0
2	0,85	0,69	0,23	0,16	0,69	19,61	22,71	0
3	0,81	0,02	0,15	0,54	0,94	9,56	12,26	0
4	0,55	0,01	0,07	3,00	0,60	8,27	33,04	0
5	0,73	0,01	0,21	0,41	0,79	16,00	22,60	0
6	0,72	0,64	0,27	0,48	1,28	20,93	30,93	0
7	0,34	0,42	0,05	1,35	0,51	8,94	20,98	1
8	0,90	0,01	0,05	0,11	0,38	3,32	3,68	0
9	0,18	0,19	0,02	5,04	0,78	0,57	3,46	0
10	0,60	0,58	0,19	0,69	1,01	13,55	22,84	0
.
.
.
220*	0,38	0,09	0,06	12,39	0,83	3,14	7,01	1
221*	0,77	0,48	0,02	0,16	0,87	0,23	0,27	1
222*	0,55	0,28	0,01	0,90	1,15	0,377	0,76	1
223*	0,44	0,16	0,04	13,78	1,20	0,223	0,536	1
224*	0,59	1,47	0,01	0,14	0,28	1,68	1,92	1
Ket: rasio likuiditas (x1), rasio profitabilitas (x2), rasio rentabilitas (x3) Rasio solvabilitas(x4),rasio aktivitas (x5),ROI (x6), ROE (x7)								

*) sintetik data

Dengan menggunakan cara yang sama, diperoleh hasil *preprocessing* data untuk tahun 2012-2013 yaitu sebagai berikut.



Gambar 4.3 Persentase Masing-masing Kelas *Financial Distress* dan *Non-Financial Distress* untuk Tahun 2012-2013

Berdasarkan Gambar 4.3 maka kelas minor merupakan kategori 1 yaitu perusahaan yang mengalami *financial distress* dengan presentase sebesar 17%. Tahapan yang dilakukan pada algoritma SMOTE adalah Setiap data pada kelas minor akan direplikasi dengan mencari tetangga terdekat (x_{knn}) menggunakan jarak *euclidean*. Dari perhitungan pada Lampiran 1.3 maka dapat diambil dua jarak *Euclidean* yang terdekat pada kelas 1 ($Y=1$ “terjadi *financial distress*”) yaitu $\sqrt{4}$ dan $\sqrt{5}$ maka kelas 1 akan direplikasi empat kali. Jumlah data kelas 1 yang semula berjumlah 22 maka setelah direplikasi sebanyak 4 kali akan menjadi 110 data. Hasil distribusi data simulasi SMOTE, dapat dilihat pada Tabel 4.7 dan data simulasi setelah SMOTE, dapat dilihat pada Tabel 4.8.

Tabel 4.7 Distribusi Data Simulasi Sebelum dan Setelah SMOTE Tahun 2012-2013

Kelas mayor	Kelas minor	Replikasi	Kelas mayor	Kelas minor baru
110(83%)	22 (17%)	4	110(50%)	110(50%)

Tabel 4.8 Data Simulasi Setelah Menggunakan SMOTE
Tahun 2012-2013

Data ke-	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆	X ₇	Y
1	0,589	0,160	0,090	0,450	0,650	11,350	16,410	0
2	0,853	0,654	0,274	0,170	0,760	20,930	24,520	0
3	0,780	0,672	0,169	0,270	1,038	11,130	14,110	0
4	0,250	0,024	0,091	2,970	0,700	7,470	29,680	0
5	0,708	0,600	0,225	0,460	0,796	18,530	27,110	0
6	0,677	0,586	0,279	0,560	1,249	16,930	26,240	0
7	0,426	0,400	0,054	1,040	0,439	7,820	15,940	1
8	0,901	0,068	0,047	0,090	0,401	3,310	3,600	0
9	0,189	0,433	0,004	4,300	0,698	5,560	29,490	0
10	0,590	0,557	0,221	0,700	1,036	15,540	26,350	0
.
.
.
216*	0,718	0,158	0,031	0,650	0,547	2,874	4,176	1
217*	0,277	0,188	0,061	2,600	1,420	4,660	16,810	1
218*	0,315	0,302	0,040	2,194	1,680	5,206	16,544	1
219*	0,353	0,084	0,065	1,893	1,321	5,493	15,655	1
220*	0,063	0,384	0,079	1,501	1,015	6,242	15,333	1

Ket: rasio likuiditas (X₁), rasio profitabilitas (X₂), rasio rentabilitas (X₃)
Rasio solvabilitas(X₄),rasio aktivitas (X₅),ROI (X₆), ROE (X₇)

*) sintetik data

4.3 Regresi Logistik Biner Data *Traininig* 2012-2013

4.3.1 Deteksi Multikolinearitas

Regresi logistik biner dilakukan untuk mengetahui faktor-faktor apa saja yang mempengaruhi kondisi *financial distress* pada perusahaan manufaktur yang tercatat di BEI . Pertama-tama dilihat nilai korelasi antar variabel prediktor apakah memiliki nilai korelasi yang tinggi antar variabel prediktor yang kemudian dilihat nilai VIF untuk menduga apakah terjadi kasus

multikolinearitas. Nilai korelasi antar variabel prediktor disajikan dalam Tabel 4.9.

Tabel 4.9 Tabel Korelasi antar Variabel Prediktor Tahun 2012-2013

	X ₁	X ₂	X ₃	X ₄	X ₅	X ₆
X ₁						
X ₂	-0,692					
	0,000					
X ₃	0,151	0,139				
	0,026	0,039				
X ₄	-0,597	0,171	-0,244			
	0,000	0,011	0,000			
X ₅	0,011	-0,127	0,120	-0,020		
	0,875	0,061	0,076	0,770		
X ₆	0,170	0,088	0,742	-0,204	-0,110	
	0,012	0,195	0,000	0,002	0,102	
X ₇	0,154	-0,004	0,666	-0,091	-0,074	0,896
	0,022	0,958	0,000	0,178	0,276	0,000
Ket: rasio likuiditas (X ₁), rasio profitabilitas (X ₂), rasio rentabilitas (X ₃) Rasio solvabilitas(X ₄),rasio aktivitas (X ₅),ROI (X ₆), ROE (X ₇)						

Berdasarkan Tabel 4.9 dapat diketahui bahwa terdapat korelasi yang tinggi antar variabel, hal tersebut dapat mengindikasikan adanya kasus multikolinearitas dan dapat menyebabkan pengujian parameter tidak signifikan. Untuk itu dilakukan deteksi multikolinearitas untuk mengetahui apakah terjadi kasus multikolinearitas dengan melihat nilai VIF yang ditunjukkan pada Tabel 4.10

Tabel 4.10 Nilai VIF Masing-masing Variabel Prediktor

No.	Variabel Prediktor	VIF	No.	Variabel Prediktor	VIF
1	Rasio Likuiditas (X ₁)	3,844	5	Rasio Aktivitas (X ₅)	1,175
2	Rasio Profitabilitas (X ₂)	2,783	6	ROI (X ₆)	7,046
3	Rasio Rentabilitas (X ₃)	2,643	7	ROE (X ₇)	5,634
4	Rasio Solvabilitas (X ₄)	1,972			

Berdasarkan Tabel 4.10 diatas dapat diketahui bahwa nilai VIF pada semua variabel telah kurang dari 10, maka dapat dikatan tidak ada indikasi kasus multikolinearitas.

4.3.2 Pengujian Parameter

Uji serentak dilakukan untuk mengetahui apakah variabel variabel X_1 - X_7 signifikan secara serentak terhadap *financial distress* pada perusahaan. Berikut pengujiannya.

Hipotesis

$$H_0 : \beta_1 = \beta_2 = \dots = \beta_p = 0$$

$$H_1 : \text{minimal terdapat satu } \beta_i \neq 0; \quad i = 1, 2, \dots, p$$

Daerah kritis

Tolak H_0 jika $\alpha < 0.05$

Statistik Uji

Tabel 4.11 Uji Serentak Tahun 2012-2013

		Chi-square	df	Sig.
Step 2	Step	12,947	1	0,000
	Block	23,259	2	0,000
	Model	23,259	2	0,000

Berdasarkan Tabel 4.11 didapatkan nilai *P-value* pada model sebesar 0,000, jika dibandingkan dengan nilai α sebesar 0,05 maka diperoleh keputusan tolak H_0 yang artinya terdapat minimal satu variabel prediktor yang berpengaruh terhadap kondisi *financial distress* perusahaan manufaktur.

Uji parsial dilakukan untuk mengetahui apakah variabel X_1 - X_7 signifikan secara parsial terhadap *financial distress* pada perusahaan. Berikut pengujiannya.

Hipotesis

$$H_0 : \beta_i = 0$$

$$H_1 : \beta_i \neq 0 \quad ; \quad i = 1, 2, \dots, p$$

Daerah kritis

Tolak H_0 jika $\alpha < 0.05$

Statistik Uji

Tabel 4.12 Uji Parsial tahun 2012-2013

	Variabel	B	S.E.	Wald	df	Sig.	Exp(B)
	X ₁	-1,925	0,609	9,999	1	0,002	0,146
Step 2 ^b	X ₅	-1,842	0,586	9,893	1	0,002	0,158
	Constant	0,925	0,659	1,972	1	0,160	2,522
Ket: rasio likuiditas (X ₁), rasio aktivitas (X ₅)							

Berdasarkan Tabel 4.12 telah dilakuakn uji signifikansi parameter dengan menggunakan subset dan diperoleh model terbaik dimana variabel yang berpengaruh signifikan terhadap kondisi *financial distress* perusahaan manufaktur adalah variabel rasio likuiditas (X₁) dan rasio aktivitas (X₅). Hal ini dapat dilihat dari *p-value* kedua variabel yang kurang dari $\alpha = 5\%$. Berikut model yang diperoleh dari Tabel 4.12.

$$\pi(x) = \frac{\exp(0,925 - 1,925x_1 - 1,842x_5)}{1 + \exp(0,925 - 1,925x_1 - 1,842x_5)}$$

Setelah mendapatkan model regresi logistik biner diatas, maka model tersebut dapat digunakan sebagai *testing* model untuk tahun 2013-2014. Untuk mengetahui perbedaan data *testing* dan data *traininig* maka dapat dilihat ketepatan klasifikasi pada sub bab berikutnya.

4.3.3 Ketepatan Klasifikasi

Berikut merupakan ketepatan klasifikasi data *training* untuk regresi logistik biner yang ditampilkan pada Tabel 4.13.

Tabel 4.13 Ketepatan Klasifikasi Data *Training* tahun 2012-2013

		Predicted		
		y		Percentage Correct
		0	1	
Step 2	Observed			
	y 0	108	2	98,2
	1	18	4	18,2
Overall Percentage				84,8

Berdasarkan Tabel 4.13 dapat diperoleh informasi jumlah data yang diklasifikasikan dalam grup yang benar dan jumlah data yang diklasifikasikan ke grup yang salah. Dapat dilihat proporsi untuk kebenaran pengklasifikasian pada kategori 0 adalah sebesar 98,2%, sedangkan proporsi untuk kebenaran pengklasifikasian pada kategori 1 adalah sebesar 18,2%, Sehingga disimpulkan proporsi benar (*accuracy rate*) keseluruhan adalah 84,8%.

Tabel 4.14 Ketepatan Klasifikasi Data *Traininig* dan Data *Testing*

		Predicted					
		Data <i>Traininig</i>			Data <i>Testing</i>		
		y		Percentage Correct	y		Percentage Correct
Observed		0	1		0	1	
Step 2 Y	0	108	2	98,2	107	2	98,2
	1	18	4	18,2	19	4	17,4
Overall Percentage				84,8	84,1		

Berdasarkan Tabel 4.14 dapat diperoleh informasi jumlah data yang diklasifikasikan dalam grup yang benar dan jumlah data yang diklasifikasikan ke grup yang salah. Dapat dilihat proporsi untuk kebenaran pengklasifikasian pada data *traininig* kategori 0 adalah sebesar 98,2%, sedangkan proporsi untuk kebenaran pengklasifikasian pada kategori 1 adalah sebesar 18,2%, Sehingga disimpulkan proporsi benar (*accuracy rate*) keseluruhan adalah 84,8%. Sedangkan Dapat dilihat proporsi untuk kebenaran pengklasifikasian pada data *testing* kategori 0 adalah sebesar 98,2%, sedangkan proporsi untuk kebenaran pengklasifikasian pada kategori 1 adalah sebesar 17,4%, Sehingga disimpulkan proporsi benar (*accuracy rate*) keseluruhan adalah 84,1%.

4.4 Klasifikasi *Financial distress* dengan Metode *Classification and Regression Tree* (CART)

Berdasarkan pada tujuan penelitian ini yaitu untuk memperoleh informasi yang berkaitan dengan variabel prediktor mana yang menjadi faktor dominan (penting) dalam menentukan klasifikasi *financial distress* perusahaan manufaktur, maka

dilakukan analisis klasifikasi dengan menggunakan pendekatan *classification and regression tree* (CART).

Sesuai pada algoritma yang telah dijelaskan pada bab dua, maka tahapan-tahapan yang dilakukan yaitu pembentukan pohon klasifikasi, dilanjutkan dengan pemangkasan pohon klasifikasi, dan terakhir yaitu penentuan pohon klasifikasi optimum.

4.4.1 Pembentukan Pohon Klasifikasi

Dalam pembentukan pohon klasifikasi terdapat beberapa langkah yaitu pemilihan pemilah, penentuan simpul termina, dan penandaan label kelas. Pemilihan pemilah setiap variabel prediktor akan dihitung ukuran keheterogenan simpul. Pemilah yang terpilih adalah variabel pemilah yang memiliki nilai *Goodness of Split* tertinggi dan merupakan variabel yang terpenting dalam mengklasifikasikan data pengamatan, seperti yang ditampilkan pada Tabel 4.15

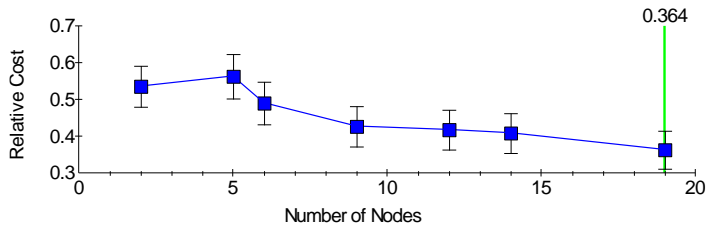
Tabel 4.15 Variabel Penting Pembentukan Pohon Klasifikasi Maksimum

Variable	Score
Rasio Likuiditas (X_1)	100,00
Rasio Rentabilitas (X_3)	98,66
Rasio Solvabilitas (X_4)	94,49
ROI (X_6)	92,43
Rasio Aktivitas (X_5)	72,26
Rasio Profitabilitas (X_2)	49,91
ROE (X_7)	49,74

Berdasarkan Tabel 4.15, dapat diperoleh informasi bahwa semua variabel prediktor menjadi pembangun dalam pembentukan pohon klasifikasi. Namun, berdasarkan skor variabel yang dihasilkan diketahui bahwa variabel X_1 mempunyai skor variabel tertinggi yaitu 100,00. Sehingga dapat dikatakan bahwa variabel rasio likuiditas (X_1) merupakan faktor terpenting dalam mengklasifikasikan *financial distress* perusahaan manufaktur. Selain itu, terdapat beberapa variabel lain yang juga berpengaruh besar dalam pembentukan pohon klasifikasi yaitu

rasio rentabilitas (X_3), rasio solvabilitas (X_4), ROI (X_6) dan rasio aktivitas (X_5).

Selanjutnya proses pemilahan terjadi sampai terbentuk pohon klasifikasi maksimum, dimana sifat dari pohon klasifikasi maksimum adalah menghasilkan simpul terminal terbanyak dengan kedalaman tertinggi. Dalam kasus ini, pohon klasifikasi sudah optimum ditunjukkan pada gambar *relative cost* yang sudah berwarna hijau ditunjukkan pada Gambar 4.4.



Gambar 4.4 Plot *Relative Cost* dan Banyaknya Simpul Terminal

Jika dilihat pada *plot relative cost* pohon klasifikasi dengan simpul terminal sebanyak 19 simpul sudah dapat dikatakan optimum ditunjukkan dengan nilai *relative cost* yang dihasilkan sebesar 0,364, ditandai dengan garis berwarna hijau. Sedangkan, untuk biaya kesalahan (*cross-validation relative cost*) menghasilkan nilai yang paling minimum yakni sebesar $0,364 \pm 0,064$ atau bekisar antara 0,400 sampai 0,300 dengan nilai kompleksitas sebesar 0,000. Berikut ini akan ditunjukkan urutan dalam pembentukan pohon klasifikasi sampai didapatkan pohon klasifikasi optimum yang ditampilkan pada Tabel 4.16 berikut.

Tabel 4.16 Urutan Pembentukan Pohon Klasifikasi (*Tree Sequence*)

<i>Tree Number</i>	<i>Terminal Nodes</i>	<i>Cross-Validated Relative Cost</i>	<i>Resubstitution Relative Cost</i>	<i>Complexity</i>
1**	19	0,364+/-0,052	0,064	0,000
2	14	0,409+/-0,054	0,109	0,005
3	12	0,418+/-0,054	0,145	0,009
4	9	0,427+/-0,055	0,218	0,012

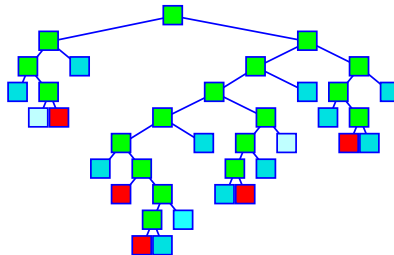
Tabel 4.16 Urutan Pembentukan Pohon Klasifikasi (*Tree Sequence*)
(lanjutan)

<i>Tree Number</i>	<i>Terminal Nodes</i>	<i>Cross-Validated Relative Cost</i>	<i>Resubstitution Relative Cost</i>	<i>Complexity</i>
5	6	0,491+/-0,058	0,3	0,014
6	5	0,564+/-0,061	0,345	0,023
7	2	0,536+/-0,056	0,509	0,027
8	1	1+/-0	1	0,245

Berdasarkan Tabel 4.16, jika dilihat pada biaya kesalahan (*cross-validation relative cost*) yang dihasilkan, pohon klasifikasi 1 memiliki biaya kesalahan yang paling minimum, sehingga pohon klasifikasi dengan simpul terminal sebanyak 19 simpul sudah layak atau dapat dikatakan optimum tanpa harus dilakukan pemangkasan terlebih dahulu.

4.4.2 Penentuan Pohon Klasifikasi Optimum

Pada sub bab pembentukan pohon klasifikasi, telah diketahui bahwa pohon klasifikasi maksimum sama dengan pohon klasifikasi optimum. Oleh sebab itu, tidak terjadi pemangkasan pohon (*pruning*), sehingga menghasilkan simpul sebanyak 19 simpul dengan tingkat kedalaman yang dihasilkan sebesar 10.



Gambar 4.5 Topologi Pohon Klasifikasi Optimum dengan Pemilahan Pemilah Indeks Gini

Berdasarkan Gambar 4.5 terdapat warna simpul dengan degradasi warna merah, biru, dan biru muda. Warna-warna tersebut menunjukkan jumlah simpul terminal sebanyak 19

simpul. Perbedaan warna tersebut berkaitan dengan pemberian label kelas masing-masing simpul terminal. Warna merah pada simpul tersebut merupakan suatu simpul yang diklasifikasikan sebagai perusahaan manufaktur dengan minimal terdapat satu perusahaan manufaktur yang mengalami kondisi *financial distress* (kelas 0). Sedangkan untuk warna biru, yaitu suatu simpul yang diklasifikasikan sebagai perusahaan manufaktur yang tidak mengalami kondisi *financial distress* (kelas 1).

Berdasarkan topologi pohon klasifikasi optimum, diketahui bahwa variabel rasio solvabilitas (X_4) merupakan variabel pemilah yang utama dan paling penting dalam klasifikasi *financial distress* perusahaan manufaktur. Selain itu, ada enam variabel lain yang juga berperan dalam pembentukan pohon klasifikasi optimum, hasil selengkapnya disajikan pada Tabel 4.17

Tabel 4.17 Variabel Penting Pembentukan Pohon Klasifikasi Optimum

Variable	Score	
Rasio Solvabilitas (X_4)	10000	
Rasio Aktivitas (X_5)	84,01	
Rasio Profitabilitas (X_2)	55,41	
Rasio Rentabilitas (X_3)	40,90	
ROI (X_6)	40,53	
Rasio Likuiditas (X_1)	10,60	
ROE (X_7)	0,00	

Variabel rasio solvabilitas (X_4) memilah simpul utama (simpul 1) menjadi simpul kiri dan simpul kanan. Simpul akan memilah kekiri jika variabel rasio solvabilitas (X_4) kurang dari sama dengan 1,020 (simpul 2) dan kekanan jika sebaliknya (simpul 5). Total sebanyak 74 data kurang dari sama dengan 1,020 menjadi simpul kiri (simpul 2) dan 146 lebih besar dari 1,020 menjadi anggota simpul kanan (simpul 5). Simpul 2 yang beranggotakan 74 perusahaan manufaktur dengan nilai rasio solvabilitas kurang dari sama dengan 1,020, selanjutnya dipilah menjadi simpul baru kiri dan kanan menurut rasio aktivitas (X_5). Jika nilai rasio aktivitas kurang dari sama dengan 0,599 maka akan dipilah menjadi anggota simpul kiri baru (simpul 3), tetapi jika nilai rasio aktivitas lebih besar dari 0,599 maka perusahaan

manufaktur tersebut akan dipilah menjadi simpul kanan baru (simpul terminal 4). Diantara 74 perusahaan manufaktur yang menjadi anggota simpul 2, diperoleh hasil bahwa terdapat 15 perusahaan manufaktur menjadi anggota simpul 3 dengan karakteristik nilai rasio solvabilitas (X_4) kurang dari sama dengan 1,020 dan nilai rasio aktivitas kurang dari sama dengan 0,599. Sisanya sebanyak 59 perusahaan manufaktur yang menjadi anggota simpul terminal 4 dengan karakteristik nilai rasio solvabilitas (X_4) kurang dari sama dengan 1,020 dan nilai rasio aktivitas lebih besar 0,599.

Simpul 3 yang beranggotakan 15 perusahaan manufaktur dengan karakteristik nilai rasio solvabilitas (X_4) kurang dari sama dengan 1,020 dan nilai rasio aktivitas kurang dari sama dengan 0,599, selanjutnya dipilah menjadi simpul baru kiri dan kanan menurut rasio profitabilitas (X_2). Jika nilai rasio profitabilitas (X_2) kurang dari sama dengan 0,155, maka perusahaan manufaktur tersebut akan dipilah ke simpul baru kiri (simpul terminal 1). Sedangkan jika nilai rasio profitabilitas (X_2) lebih besar 0,155, maka akan dipilah ke simpul kanan baru (simpul 4).

Simpul akan terus dipilah menjadi simpul anak baru yaitu simpul kiri dan simpul kanan sesuai prosedur *binary recursive partitioning* sampai simpul tersebut telah dianggap memiliki anggota yang homogen maka simpul akan menjadi simpul terminal dan tidak akan dipilah lagi. Gambar 4.6 merupakan visualisasi struktur pohon klasifikasi optimum berdasarkan penjelasan pemilihan simpul yang telah dijelaskan diatas.

Berikut merupakan rangkuman pengklasifikasian kelas perusahaan manufaktur menurut indikasi kesamaan label kelas setiap simpul terminal yang disajikan pada Tabel 4.18.

Tabel 4.18 Kelas Perusahaan Manufaktur pada Masing-masing Simpul Terminal

Kelas	Simpul Terminal ke-	Perse ntase	Simpul Terminal ke-	Perse ntase
0 (perusahaan manufaktur tidak mengalami <i>financial distress</i>)	1	100	11	100
	2	66,7	13	100

Tabel 4.18 Kelas Perusahaan Manufaktur pada Masing-masing Simpul Terminal (lanjutan)

Kelas	Simpul Terminal ke-	Perse ntase	Simpul Terminal ke-	Perse ntase
0 (perusahaan manufaktur tidak mengalami <i>financial distress</i>)	4	100	14	71,4
	5	100	15	100
	8	100	16	100
	9	85,7	18	100
	10	100	19	100
1(Perusahaan Manufaktur Mengalami <i>Financial distress</i>)	3	90	12	96,3
	6	100	17	98,3
	7	100		

Karakteristik pada masing-masing kelas pada variabel respon dapat diketahui dengan menelusuri pohon klasifikasi optimum yang sudah terbentuk berdasarkan aturan pemilahan indeks gini. Karakteristik yang kuat didapatkan pada simpul terminal yang memiliki persentase kelas terbesar (100%) pada pelabelan kelas. Berikut adalah karakteristik masing-masing kelas yang disajikan pada Tabel 4.19.

Tabel 4.19 Karakteristik Kelas Perusahaan Manufaktur Menurut Persentase Kelas Terbesar

Kelas	Karakteristik
0 (perusahaan manufaktur tidak mengalami <i>financial distress</i>)	Rasio aktivitas $\leq 0,599$, dan rasio solvabilitas $\leq 1,020$
(1) Perusahaan Manufaktur Mengalami <i>Financial distress</i>	Rasio likuiditas $\leq 0,634$, rasio profitabilitas $\leq 0,155$, rasio aktivitas $\leq 0,599$, dan rasio solvabilitas $\leq 1,020$

4.4.3 Ketepatan Klasifikasi Pohon CART

Hasil klasifikasi *financial distress* perusahaan manufaktur berdasarkan masing-masing kelas diukur berdasarkan total ketepatan klasifikasi, akurasi pada masing-masing fold, dan total kesalahan klasifikasi untuk data *traininig* dan data *testing*. Hasil klasifikasi berdasarkan data *learning* sebelum dilakukan *pre-processing* dengan SMOTE dapat ditunjukkan pada Tabel 4.20 berikut ini.

Tabel 4.20 Klasifikasi Data *Traininig* pada Pohon Klasifikasi Optimum Sebelum SMOTE

Fold	Akurasi tiap kelas		Total akurasi
	Kelas 0	Kelas 1	
1	1	0,8000	0,9576
2	1	0,5758	0,8814
3	1	0,7143	0,9328
4	1	0,6667	0,9160
5	1	0,7308	0,9412
6	1	0,7407	0,9412
7	1	0,7143	0,9328
8	1	0,7143	0,9328
9	1	0,6250	0,8992
10	1	0,6897	0,9244
Mean	1	0,6971	0,9259

Ket: kelas 0 (perusahaan tidak mengalami *financial distress*), dan kelas 1 (perusahaan mengalami *financial distress*)

Berdasarkan Tabel 4.20, dapat diketahui rata-rata total akurasi data *traininig* dengan metode pemilah indeks gini sebelum dilakukan *pre-processing* dengan SMOTE adalah sebesar 92,6%. Namun, jika dilihat dari akurasi untuk masing-masing kelas akurasi yang dihasilkan saling berbeda jauh, seperti akurasi yang dihasilkan untuk kelas 0 mencapai 100%. Sedangkan, untuk kelas 1 nilai akurasinya hanya 69,7%, hal ini menunjukkan bahwa akurasi yang dihasilkan oleh metode 10-*fold cross validation* belum cukup stabil. Untuk mengetahui seberapa layak pohon klasifikasi yang terbentuk maka dapat diuji dengan menggunakan data *testing* sebelum dilakukan *pre-processing* dengan menggunakan metode SMOTE sebagai berikut.

Tabel 4.21 Ketepatan Klasifikasi Data *Testing* Sebelum SMOTE

Kelas Observasi	Kelas Aktual		Total	Akurasi	APER	1-APER
	Kelas 0	Kelas 1				
Kelas 0	64	46	110	0,5818	0,3864	0,6136
Kelas 1	5	17	21	0,7727		
Total	69	63	132			

Berdasarkan Tabel 4.21, dapat diketahui total akurasi data *testing* dengan metode pemilah indeks gini sebelum dilakukan

pre-processing dengan menggunakan metode SMOTE adalah sebesar 61,36%. Diketahui bahwa nilai total akurasi yang dihasilkan dengan menggunakan indeks gini masih tergolong kecil. Hal tersebut disebabkan oleh jumlah data untuk masing-masing kategori pada variabel respon tidak seimbang sehingga terjadi kasus *underfitting*. Sehingga pohon klasifikasi yang sudah terbentuk dengan menggunakan data *traininig* kurang layak untuk digunakan dalam pengklasifikasian *financial distress* perusahaan manufaktur. Oleh karena itu, diperlukan suatu metode *pre-processing* dengan menggunakan metode SMOTE untuk mendapatkan total akurasi yang lebih tinggi. Berikut adalah hasil ketepatan klasifikasi *financial distress* perusahaan manufaktur untuk data *traininig* dan *testing* setelah dilakukan *pre-processing* dengan menggunakan SMOTE.

Tabel 4.22 Klasifikasi Data *Traininig* pada Pohon Klasifikasi Optimum Sesudah SMOTE

Fold	Akurasi tiap kelas		Total akurasi
	Kelas 0	Kelas 1	
1	0,9505	0,9691	0,9596
2	0,9608	0,9896	0,9747
3	0,9894	0,9423	0,9646
4	1	0,9340	0,9646
5	0,9794	0,9604	0,9697
6	0,9691	0,9505	0,9596
7	0,9694	0,9600	0,9646
8	0,9583	0,9314	0,9444
9	0,9694	0,9600	0,9646
10	0,9600	0,9694	0,9646
<i>Mean</i>	0,9706	0,9567	0,9631
Ket: kelas 0 (perusahaan tidak mengalami <i>financial distress</i>), dan kelas 1 (perusahaan mengalami <i>financial distress</i>)			

Pada Tabel 4.20 dapat diperoleh informasi dapat diketahui rata-rata akurasi data *trainig* dengan metode pemilah indeks gini sesudah dilakukan *pre-processing* dengan SMOTE

adalah sebesar 96,3%. Untuk mengetahui seberapa layak pohon klasifikasi yang terbentuk maka dapat diuji dengan menggunakan data *testing* sesudah dilakukan *pre-processing* dengan menggunakan metode SMOTE sebagai berikut.

Tabel 4.23 Ketepatan Klasifikasi Data *Testing* pada Pohon Klasifikasi Optimum Sesudah SMOTE

Kelas Observasi	Kelas Aktual		Total	Akurasi	APER	1-APER
	Kelas 0	Kelas 1				
Kelas 0	85	24	109	0,7798	0,1786	0,8214
Kelas 1	16	94	115	0,8174		
Total	102	123	224			

Pada Tabel 4.23 dapat diketahui akurasi data *testing* dengan metode pemilah indeks gini sebelum dilakukan *pre-processing* dengan menggunakan metode SMOTE adalah sebesar 77,9% dengan keterangan perusahaan manufaktur yang termasuk kategori mengalami *financial distress* (kelas 1) yang tepat diklasifikasikan sebagai perusahaan yang mengalami *financial distress* adalah sebanyak 85 perusahaan, sedangkan sebanyak 24 perusahaan yang mengalami *financial distress* salah diklasifikasikan menjadi perusahaan yang tidak mengalami *financial distress*. Akurasi untuk kelas tidak mengalami *financial distress* (kelas 0) yang dihasilkan adalah sebesar 81,8% dengan ketentuan perusahaan manufaktur yang tidak mengalami *financial distress* (kelas 0) yang tepat diklasifikasikan adalah sebanyak 94 perusahaan, sedangkan sebanyak 16 perusahaan mengalami kesalahan pengklasifikasian menjadi perusahaan yang mengalami *financial distress*.

Berdasarkan akurasi pada masing-masing kelas atau kategori pada variabel respon didapatkan total akurasi (1-APER) yang dihasilkan untuk klasifikasi *financial distress* perusahaan manufaktur pada tahun 2012-2014 dengan menggunakan pemilihan pemilah indeks gini adalah sebesar 82,1%. Sedangkan, total kesalahan klasifikasi (APER) yang dihasilkan adalah sebesar 1,78%. Akurasi klasifikasi *financial distress* perusahaan manufaktur setelah dilakukan *pre-processing* dengan SMOTE

lebih tinggi, sehingga dapat dikatakan bahwa pohon klasifikasi optimal yang dihasilkan dengan jumlah simpul terminal sebanyak 19 simpul sudah layak, sehingga pohon klasifikasi tersebut dapat digunakan untuk mengklasifikasikan *financial distress* perusahaan manufaktur dengan menggunakan data yang baru. Berikut adalah perbandingan hasil ketepatan klasifikasi pohon optimum data *traininig* dan data *testing* setelah dilakukan *pre-processing* dengan SMOTE yang ditunjukkan pada Tabel 4.24.

Tabel 4.24 Perbandingan Ketepatan Klasifikasi Pohon Optimum Sebelum dan Sesudah SMOTE

Pohon Klasifikasi	Ketepatan Klasifikasi (%)	
	<i>Training</i>	<i>Testing</i>
Optimal		
Sebelum SMOTE	93,3	61,3
Sesudah SMOTE	96,3	82,1

Berdasarkan Tabel 4.24 diatas, dapat diperoleh informasi bahwa nilai ketepatan klasifikasi sesudah dilakukan SMOTE menghasilkan 3% lebih besar dari pada sebelum dilakukan SMOTE untuk data *training*. Sedangkan untuk data *testing* dihasilkan nilai ketepatan klasifikasi 20,8% lebih besar setelah dilakukan SMOTE.

4.5 Perbandingan Hasil Klasifikasi Antara Metode Regresi Logistik Biner dan CART

Berdasarkan pada tujuan penelitian ini yaitu untuk memperoleh informasi yang berkaitan dengan klasifikasi dengan dua pendekatan parametrik dan nonparametrik, maka digunakan analisis regresi logistik biner untuk pendekatan parametrik, dan *classification analysis and regression tree* (CART) untuk pendekatan noon parametrik. Berikut merupakan perbandingan hasil ketepatan klasifikasi antara dua metode tersebut yang ditampilkan pada Tabel 4.25.

Tabel 4.25 Perbandingan Ketepatan Klasifikasi Antara Metode Regresi Logistik dan Metode CART

Metode	Ketepatan Klasifikasi (%)	
	<i>Traininig</i>	<i>Testing</i>
RegLog Biner	84,8	84,1
CART	96,3	82,1

Menurut Hosmer dan Lemeshow (1989) salah satu ukuran kebaikan model adalah jika memiliki peluang kesalahan klasifikasi yang minimal dan ketepatan prediksi dari model yang maksimum. Berdasarkan Tabel 4.25 untuk data *traininig* diperoleh ketepatan klasifikasi analisis regresi logistik sebesar 84,8% dan ketepatan klasifikasi metode CART sebesar 96,3%. Hal ini menunjukkan bahwa metode CART memiliki nilai ketepatan klasifikasi yang lebih besar 11,5% dari metode CART. Akan tetapi nilai klasifikasi pada metode regresi logistik untuk data *testing* 2% lebih bagus dibandingkan dengan metode CART.

Perbedaan tingkat ketepatan prediksi dapat disebabkan oleh perbedaan hasil pengklasifikasian. Dalam regresi logistik variabel prediktor yang terpilih sebagai variabel yang berpengaruh terhadap variabel respon adalah rasio likuiditas (X_1) dan rasio aktivitas (X_5). Sedangkan pada metode CART variabel yang terpilih adalah rasio profitabilitas (X_2), rasio rentabilitas (X_3), rasio solvabilitas (X_4), rasio aktivitas (X_5), dan ROI (X_6). Variabel rasio aktivitas (X_5) merupakan variabel yang konsisten berpengaruh pada analisis regresi logistik maupun metode CART.

LAMPIRAN

Lampiran A. Data Sebelum SMOTE Tahun 2013-2014

y	x1	x2	x3	x4	x5	x6	x7
0	0.509	0.027	0.059	0.700	0.612	6.760	11.470
0	0.858	0.698	0.235	0.160	0.692	19.610	22.710
0	0.813	0.019	0.153	0.540	0.937	9.560	12.260
0	0.552	0.013	0.074	3.000	0.605	8.270	33.040
0	0.729	0.007	0.207	0.410	0.786	16.000	22.600
0	0.724	0.641	0.277	0.480	1.278	20.930	30.930
1	0.345	0.423	0.051	1.350	0.506	8.940	20.980
0	0.900	0.002	0.050	0.110	0.382	3.320	3.680
0	0.183	0.186	0.023	5.040	0.780	0.570	3.460
0	0.607	0.575	0.188	0.690	1.013	13.550	22.840
0	0.258	0.077	0.016	3.050	5.024	1.950	7.910
0	0.200	0.089	0.001	3.190	1.038	0.950	3.970
....
....
....
....
1	0.501	1.135	0.018	1.180	0.252	10.480	22.810
0	0.757	0.443	0.179	0.250	1.293	11.240	14.100
0	0.540	0.138	0.019	0.910	0.769	11.800	22.510
0	0.505	0.209	0.029	2.090	1.315	1.010	3.640
0	0.525	0.510	0.101	0.580	1.397	16.080	25.410
0	0.305	0.230	0.031	1.570	1.275	10.910	28.020
0	0.459	0.042	0.009	1.460	0.895	0.270	0.660
1	-0.407	2.314	0.017	3.500	0.590	14.590	17.520
1	0.115	0.190	0.034	1.210	1.141	19.070	23.710
0	0.447	0.110	0.079	0.130	1.384	10.910	13.510

Lampiran B. Data Sebelum SMOTE Tahun 2012-2013

y	x1	x2	x3	x4	x5	x6	x7
0	0.589	0.160	0.090	0.450	0.650	11.350	16.410
0	0.853	0.654	0.274	0.170	0.760	20.930	24.520
0	0.780	0.672	0.169	0.270	1.038	11.130	14.110
0	0.250	0.024	0.091	2.970	0.700	7.470	29.680
0	0.708	0.600	0.225	0.460	0.796	18.530	27.110
0	0.677	0.586	0.279	0.560	1.249	16.930	26.240
1	0.426	0.400	0.054	1.040	0.439	7.820	15.940
0	0.901	0.068	0.047	0.090	0.401	3.310	3.600
0	0.189	0.433	0.004	4.300	0.698	5.560	29.490
0	0.590	0.557	0.221	0.700	1.036	15.540	26.350
0	0.371	0.144	0.042	1.700	5.659	4.240	11.430

Lampiran B. Data Sebelum SMOTE Tahun 2012-2013 (lanjutan)

0	0.312	0.169	0.001	2.200	1.712	0.740	2.370
0	0.780	0.653	0.223	0.280	1.068	16.990	18.470
0	0.531	0.340	0.175	0.880	0.748	12.860	24.210
....
....
....
....
0	0.732	0.662	0.256	0.370	1.633	18.930	25.870
0	0.646	0.237	0.059	0.550	1.301	3.910	6.050
0	0.890	0.147	0.197	0.650	0.804	18.020	29.700
0	0.014	0.006	0.009	1.540	0.545	4.080	14.110
0	0.713	0.186	0.098	0.400	1.178	7.600	10.670
0	0.869	0.640	0.161	0.150	1.467	11.950	13.750
0	0.847	0.605	0.093	0.180	1.006	7.560	8.920
0	0.597	0.580	0.257	0.680	2.125	21.090	35.330
0	0.701	0.060	0.032	0.430	0.998	2.380	3.390
0	0.502	0.116	0.006	1.970	0.734	11.480	12.470

Lampiran C. Data Sesudah SMOTE Tahun 2013-2014

x1	x2	x3	x4	x5	x6	x7	y
0.588986	0.159552	0.089731	0.45	0.650303	11.35	16.41	0
0.853377	0.653885	0.274204	0.17	0.759842	20.93	24.52	0
0.78	0.672049	0.168901	0.27	1.037519	11.13	14.11	0
0.250229	0.024446	0.091033	2.97	0.699743	7.47	29.68	0
0.708085	0.600168	0.22474	0.46	0.795679	18.53	27.11	0
0.676938	0.586148	0.278662	0.56	1.248753	16.93	26.24	0
0.426107	0.399594	0.054237	1.04	0.438793	7.82	15.94	1
0.901447	0.068294	0.046988	0.09	0.401094	3.31	3.6	0
0.188684	0.433402	0.003585	4.3	0.69839	5.56	29.49	0
0.589864	0.557052	0.220851	0.7	1.035529	15.54	26.35	0
0.370742	0.143611	0.042145	1.7	5.659141	4.24	11.43	0
0.312443	0.168542	0.001391	2.2	1.712207	0.74	2.37	0
0.780004	0.65333	0.22323	0.28	1.068264	16.99	18.47	0
0.531252	0.340274	0.174607	0.88	0.748245	12.86	24.21	0
....
....
....
....
0.078763	2.263535	0.065137	1.391	0.48895	5.839	4.807	1
-0.54287	1.895688	0.112902	1.923	0.861735	5.181	5.024	1

Lampiran C. Data Sesudah SMOTE Tahun 2013-2014 (lanjutan)

0.351657	0.096772	0.022605	1.84	0.684299	1.74	4.96	1
0.156181	0.393763	0.088616	1.812	0.620259	3.315	5.303	1
0.410418	0.06812	0.011009	1.462	0.830288	1.334	3.357	1
0.718081	0.158359	0.031007	0.65	0.547291	2.874	4.176	1
0.2774	0.188449	0.061764	2.6	1.42021	4.66	16.81	1
0.315142	0.302434	0.04086	2.194	1.680038	5.206	16.544	1
0.353168	0.084379	0.065522	1.893	1.32182	5.493	15.655	1
0.063451	0.384419	0.079108	1.501	1.01572	6.242	15.333	1

Lampiran D. Data Sesudah SMOTE Tahun 2012-2013

x1	x2	x3	x4	x5	x6	x7	y
0.509358	0.026738	0.058585	0.7	0.612301	6.76	11.47	0
0.858052	0.697937	0.235057	0.16	0.692272	19.61	22.71	0
0.812742	0.019013	0.152564	0.54	0.937167	9.56	12.26	0
0.55204	0.013359	0.07384	3	0.604626	8.27	33.04	0
0.728623	0.007383	0.20664	0.41	0.786458	16	22.6	0
0.724467	0.641099	0.276673	0.48	1.278423	20.93	30.93	0
0.344535	0.422612	0.050982	1.35	0.505878	8.94	20.98	1
0.899791	0.001867	0.05047	0.11	0.38213	3.32	3.68	0
0.183166	0.185988	0.02266	5.04	0.78026	0.57	3.46	0
0.60731	0.575416	0.188371	0.69	1.012994	13.55	22.84	0
0.258233	0.077139	0.016067	3.05	5.02437	1.95	7.91	0
0.199548	0.088827	0.000509	3.19	1.038491	0.95	3.97	0
0.841997	0.736757	0.055006	0.27	0.545532	14.56	18.47	0
....
....
....
....
0.45859	0.853493	0.063667	3.667	0.588618	10.723	24.418	1
0.348715	0.354999	0.08678	3.786	0.766517	9.645	23.137	1
0.474199	0.002873	0.037125	1.19	1.106616	4.19	9.18	1
0.05009	0.450405	0.076328	1.092	1.062077	3.301	9.061	1
0.39593	0.320941	0.023356	1.701	2.030657	3.399	9.012	1
0.382524	0.089491	0.060397	1.239	0.835441	3.14	7.01	1
0.769758	0.481923	0.020131	0.16	0.87161	0.23	0.27	1
0.547262	0.288099	0.015105	0.909	1.15674	0.377	0.76	1
0.436663	0.160472	0.044274	1.378	1.20072	0.223	0.536	1

Lampiran E. Output Korelasi Untuk Regresi Logistik Biner

Correlations: y; x1; x2; x3; x4; x5; x6; x7

	x1	x2	x3	x4	x5	x6
x1						
x2	-0.692 0.000					
x3	0.151 0.026	0.139 0.039				
x4	-0.597 0.000	0.171 0.011	-0.244 0.000			
x5	0.011 0.875	-0.127 0.061	0.120 0.076	-0.020 0.770		
x6	0.170 0.012	0.088 0.195	0.742 0.000	-0.204 0.002	-0.110 0.102	
x7	0.154 0.022	-0.004 0.958	0.666 0.000	-0.091 0.178	-0.074 0.276	0.896 0.000

Lampiran F. Output Deteksi Multikolinearitas

Predictor	Coef	SE Coef	T	P	VIF
Constant	0.9125	0.1286	7.09	0.000	
x1	-0.3274	0.1207	-2.71	0.007	3.844
x2	0.00909	0.07138	0.13	0.899	2.783
x3	-1.9212	0.5429	-3.54	0.000	2.643
x4	-0.00298	0.03024	-0.10	0.922	1.972
x5	-0.22010	0.05059	-4.35	0.000	1.175
x6	0.005351	0.007276	0.74	0.463	7.046
x7	0.004178	0.005523	0.76	0.450	5.634

Lampiran G. Output Pengujian Parameter Serentak dan Parsial

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	10.312	1	.001
	Block	10.312	1	.001
	Model	10.312	1	.001

Lampiran G. Output Pengujian Parameter Serentak dan Parsial (lanjutan)

Step 2	Step	12.947	1	.000
	Block	23.259	2	.000
	Model	23.259	2	.000

Variables in the Equation

	B	S.E.	Wald	df	Sig.	Exp(B)
Step 1 ^a x1	-1.677	.650	6.660	1	.010	.187
Constant	-.913	.367	6.201	1	.013	.401
Step 2 ^b x1	-1.925	.609	9.999	1	.002	.146
x5	-1.842	.586	9.893	1	.002	.158
Constant	.925	.659	1.972	1	.160	2.522

a. Variable(s) entered on step 1: x1.

b. Variable(s) entered on step 2: x5.

Lampiran H. Output Tabel Klasifikasi

Classification Table^c

Observed		Predicted					
		Selected Cases ^a			Unselected Cases ^b		
		y		Percentage Correct	y		Percentage Correct
		0	1		0	1	
Step 1	y 0	109	1	99.1	108	1	99.1
	1	20	2	9.1	21	2	8.7
	Overall Percentage			84.1			83.3
Step 2	y 0	108	2	98.2	107	2	98.2
	1	18	4	18.2	19	4	17.4
	Overall Percentage			84.8			84.1

a. Selected cases kode

EQ 0

b. Unselected cases kode

NE 0

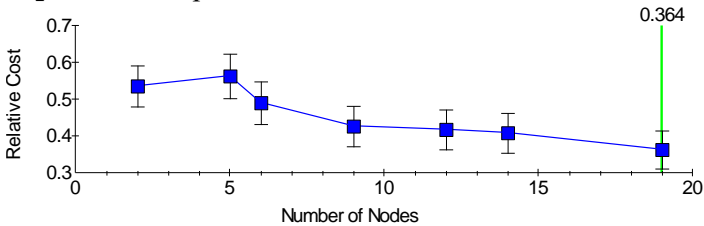
Lampiran I. Output Variabel Penting Pohon Klasifikasi Maksimum

Variable	Score	
X1	100.00	<div></div>
X3	98.66	<div></div>
X4	94.49	<div></div>
X6	92.43	<div></div>
X5	72.26	<div></div>
X2	49.91	<div></div>
X7	49.74	<div></div>

Lampiran J. Output Variabel Penting Pohon Klasifikasi Optimal

Variable	Score	
X4	100.00	<div></div>
X5	84.01	<div></div>
X2	55.41	<div></div>
X3	40.90	<div></div>
X6	40.53	<div></div>
X1	10.60	<div></div>
X7	0.00	

Lampiran K. Output Grafik Relative Cost



Lampiran L. Simpul Pohon Klasifikasi

Dependent variable: Y

Terminal Tree Nodes		Cross-Validated Relative Cost	Resubstitution Relative Cost	Complexity Parameter
1**	19	0.364 +/- 0.052	0.064	0.000
2	14	0.409 +/- 0.054	0.109	0.005
3	12	0.418 +/- 0.054	0.145	0.009
4	9	0.427 +/- 0.055	0.218	0.012
5	6	0.491 +/- 0.058	0.300	0.014
6	5	0.564 +/- 0.061	0.345	0.023
7	2	0.536 +/- 0.056	0.509	0.027
8	1	1.000 +/- 0.000	1.000	0.245

Initial misclassification cost = 0.500

Initial class assignment = 0

```
=====
NODE INFORMATION
=====
```

```
*****
*                Node 1: X4                *
*                N: 220                     *
*****
*****
*                *                *                *
*                *                *                *
*                *                *                *
*                *                *                *
*                *                *                *
```

Node 1 was split on X4
 A case goes left if X4 <= 1.020
 Improvement = 0.135 Complexity Threshold = 0.245

Node	Cases	Wgt	Counts	Cost	Class
1	220		220.00	0.500	0
2	74		74.00	0.135	0
5	146		146.00	0.315	1

Class	Weighted Counts		
	Top	Left	Right
0	110.00	64.00	46.00
1	110.00	10.00	100.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.500	0.865	0.315
1	0.500	0.135	0.685

Surrogate		Split	Assoc.	Improve.
1 X1	r	0.495	0.878	0.107
2 X3	r	0.121	0.392	0.078
3 X6	r	9.400	0.378	0.057
4 X7	r	17.575	0.095	0.026
5 X5	s	0.031	0.027	0.005

Competitor	Split	Improve.
1 X1	0.467	0.127
2 X3	0.121	0.078
3 X6	8.080	0.069
4 X5	0.731	0.064
5 X2	0.772	0.048

```

*****
*                               *
*           Node 2: X5          *
*           N: 74              *
*                               *
*****
*****
Node 3                               = Terminal Node 4
N: 15                               = N: 59
*****

```

Node 2 was split on X5

A case goes left if X5 <= 0.599

Improvement = 0.048 Complexity Threshold = 0.023

Node	Cases	Wgt	Counts	Cost	Class
2	74		74.00	0.135	0
3	15		15.00	0.333	1
-4	59		59.00	0.000	0

Class	Weighted Counts		
	Top	Left	Right
0	64.00	5.00	59.00
1	10.00	10.00	0.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.865	0.333	1.000
1	0.135	0.667	0.000

Surrogate		Split	Assoc.	Improve.
1 X1	r	0.873	0.333	0.026
2 X4	s	0.145	0.267	0.017
3 X6	r	33.488	0.200	0.021
4 X7	r	40.366	0.133	0.015
5 X3	s	0.012	0.133	.341244E-03

Competitor	Split	Improve.
1 X1	0.873	0.026
2 X6	33.488	0.021
3 X4	0.145	0.017
4 X7	40.366	0.015
5 X3	0.046	0.008

```

*****
*               Node 3: X2               *
*               N: 15                     *
*****
=====
= Terminal Node 1 = * Node 4 *
= N: 2 = * N: 13 *
=====
*****

```

Node 3 was split on X2

A case goes left if $X2 \leq 0.155$

Improvement = 0.009 Complexity Threshold = 0.009

Node	Cases	Wgt	Counts	Cost	Class
3	15		15.00	0.333	1
-1	2		2.00	0.000	0
4	13		13.00	0.231	1

Class	Weighted Counts		
	Top	Left	Right
0	5.00	2.00	3.00
1	10.00	0.00	10.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.333	1.000	0.231
1	0.667	0.000	0.769

Surrogate	Split	Assoc.	Improve.
1 X3 s	0.005	0.500	0.004

Competitor	Split	Improve.
1 X3	0.016	0.009
2 X6	1.265	0.009
3 X7	1.513	0.009
4 X1	0.634	0.009
5 X5	0.517	0.004

```

*****
*                               *
*           Node 1: X4          *
*           N: 220              *
*                               *
*****
*****
*                               *
*           Node 2              *
*           N: 74               *
*                               *
*****
*                               *
*           Node 5              *
*           N: 146              *
*                               *
*****

```

Node 1 was split on X4

A case goes left if $X4 \leq 1.020$

Improvement = 0.135 Complexity Threshold = 0.245

Node	Cases	Wgt	Counts	Cost	Class
1	220		220.00	0.500	0
2	74		74.00	0.135	0
5	146		146.00	0.315	1

Class	Weighted Counts		
	Top	Left	Right
0	110.00	64.00	46.00
1	110.00	10.00	100.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.500	0.865	0.315
1	0.500	0.135	0.685

Surrogate		Split	Assoc.	Improve.
1 X1	r	0.495	0.878	0.107
2 X3	r	0.121	0.392	0.078
3 X6	r	9.400	0.378	0.057
4 X7	r	17.575	0.095	0.026
5 X5	s	0.031	0.027	0.005

Competitor	Split	Improve.
1 X1	0.467	0.127
2 X3	0.121	0.078
3 X6	8.080	0.069
4 X5	0.731	0.064
5 X2	0.772	0.048


```

*****
*                               *
*           Node 2: X5          *
*           N: 74              *
*                               *
*****
*****
*                               *
*           Node 3            *
*           N: 15             *
*                               *
*****
=====
*                               *
*           Node 3            *
*           N: 15             *
*                               *
*****
=====

```

Node 2 was split on X5

A case goes left if X5 <= 0.599

Improvement = 0.048 Complexity Threshold = 0.023

Node	Cases	Wgt	Counts	Cost	Class
2	74		74.00	0.135	0
3	15		15.00	0.333	1
-4	59		59.00	0.000	0

Class	Weighted Counts		
	Top	Left	Right
0	64.00	5.00	59.00
1	10.00	10.00	0.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.865	0.333	1.000
1	0.135	0.667	0.000

	Surrogate		Split	Assoc.	Improve.
1	X1	r	0.873	0.333	0.026
2	X4	s	0.145	0.267	0.017
3	X6	r	33.488	0.200	0.021
4	X7	r	40.366	0.133	0.015
5	X3	s	0.012	0.133	.341244E-03

	Competitor	Split	Improve.
1	X1	0.873	0.026
2	X6	33.488	0.021
3	X4	0.145	0.017
4	X7	40.366	0.015
5	X3	0.046	0.008

```

*****
*                               *
*           Node 3: X2          *
*           N: 15              *
*                               *
*****
=====
= Terminal Node 1 = * Node 4 *
= N: 2 = * N: 13 *
=====
*****

```

Node 3 was split on X2

A case goes left if X2 <= 0.155

Improvement = 0.009 Complexity Threshold = 0.009

Node	Cases	Wgt	Counts	Cost	Class
3	15		15.00	0.333	1
-1	2		2.00	0.000	0
4	13		13.00	0.231	1

Class	Weighted Counts		
	Top	Left	Right
0	5.00	2.00	3.00
1	10.00	0.00	10.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.333	1.000	0.231
1	0.667	0.000	0.769

Surrogate	Split	Assoc.	Improve.
1 X3 s	0.005	0.500	0.004

Competitor	Split	Improve.
1 X3	0.016	0.009
2 X6	1.265	0.009
3 X7	1.513	0.009
4 X1	0.634	0.009
5 X5	0.517	0.004

```

*****
*               Node 4: X1               *
*               N: 13                     *
*****
=====
=      Terminal Node 2      =      Terminal Node 3      =
=             N: 3         =             N: 10         =
=====

```

Node 4 was split on X1

A case goes left if $X1 \leq 0.634$

Improvement = 0.007 Complexity Threshold = 0.005

Node	Cases	Wgt	Counts	Cost	Class
4	13		13.00	0.231	1
-2	3		3.00	0.333	0
-3	10		10.00	0.100	1

Class	Weighted Counts		
	Top	Left	Right
0	3.00	2.00	1.00
1	10.00	1.00	9.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.231	0.667	0.100
1	0.769	0.333	0.900

	Surrogate		Split	Assoc.	Improve.
1	X3	s	0.016	0.333	0.006
2	X6	s	1.260	0.333	0.006
3	X7	s	1.493	0.333	0.006
4	X2	r	0.338	0.333	0.002

	Competitor		Split	Improve.
1	X3		0.016	0.006
2	X6		1.260	0.006
3	X7		1.493	0.006
4	X5		0.444	0.004
5	X4		0.629	0.003

```

*****
*                               *
*           Node 5: X2          *
*           N: 146              *
*                               *
*****
*****
*                               *
*           Node 6              *
*           N: 84               *
*                               *
*****
*                               *
*           Node 16             *
*           N: 62               *
*                               *
*****

```

Node 5 was split on X2

A case goes left if $X2 \leq 0.362$

Improvement = 0.054 Complexity Threshold = 0.027

Node	Cases	Wgt	Counts	Cost	Class
5	146		146.00	0.315	1
6	84		84.00	0.488	1
16	62		62.00	0.081	1

Class	Weighted Counts		
	Top	Left	Right
0	46.00	41.00	5.00
1	100.00	43.00	57.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.315	0.488	0.081
1	0.685	0.512	0.919

Surrogate		Split	Assoc.	Improve.
1 X5	r	0.623	0.435	0.006
2 X1	r	0.117	0.419	0.018
3 X3	s	0.069	0.274	.680983E-03
4 X6	s	5.875	0.210	0.001
5 X7	s	18.313	0.194	0.002

Competitor	Split	Improve.
1 X5	1.826	0.027
2 X1	0.099	0.020
3 X3	0.134	0.015
4 X6	11.070	0.011
5 X7	17.070	0.009

```

*****
*                               *
*           Node 6: X6          *
*           N: 84              *
*                               *
*****
*****
*                               *
*           Node 7              *
*           N: 72              *
*                               *
*****
*****

```

Node 6 was split on X6

A case goes left if X6 <= 6.875

Improvement = 0.033 Complexity Threshold = 0.041

Node	Cases	Wgt	Counts	Cost	Class
6	84		84.00	0.488	1
7	72		72.00	0.403	1
-15	12		12.00	0.000	0

Class	Weighted Counts		
	Top	Left	Right
0	41.00	29.00	12.00
1	43.00	43.00	0.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.488	0.403	1.000
1	0.512	0.597	0.000

	Surrogate		Split	Assoc.	Improve.
1	X3	s	0.086	0.333	0.021
2	X7	s	17.395	0.250	0.018
3	X5	r	0.338	0.167	0.005
4	X2	s	0.262	0.083	0.002

	Competitor		Split	Improve.
1	X3		0.081	0.024
2	X5		0.592	0.021
3	X7		17.070	0.021
4	X1		0.467	0.009
5	X4		1.069	0.005

```

*****
*                               *
*           Node 7: X6          *
*           N: 72              *
*                               *
*****
*****
*                               * *                               *
*           Node 8              * *           Node 13              *
*           N: 36              * *           N: 36              *
*                               * *                               *
*****

```

Node 7 was split on X6

A case goes left if X6 <= 4.625

Improvement = 0.021 Complexity Threshold = 0.027

Node	Cases	Wgt	Counts	Cost	Class
7	72		72.00	0.403	1
8	36		36.00	0.417	0
13	36		36.00	0.222	1

Class	Weighted Counts		
	Top	Left	Right
0	29.00	21.00	8.00
1	43.00	15.00	28.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.403	0.583	0.222
1	0.597	0.417	0.778

Surrogate		Split	Assoc.	Improve.
1 X3	s	0.051	0.778	0.007
2 X7	s	10.025	0.722	0.008
3 X5	s	0.904	0.528	.179306E-03
4 X2	s	0.124	0.444	.404030E-03
5 X1	r	0.350	0.167	.277117E-03

Competitor	Split	Improve.
1 X7	11.775	0.017
2 X5	0.592	0.014
3 X2	0.223	0.013
4 X3	0.081	0.010
5 X1	0.467	0.008


```

*****
*                               *
*           Node 9: X5          *
*           N: 28              *
*                               *
*****
=====
= Terminal Node 5 = *           Node 10          *
= N: 4           = *           N: 24           *
=====
*****

```

Node 9 was split on X5

A case goes left if X5 <= 0.592

Improvement = 0.012 Complexity Threshold = 0.018

Node	Cases	Wgt	Counts	Cost	Class
9	28		28.00	0.464	1
-5	4		4.00	0.000	0
10	24		24.00	0.375	1

Weighted Counts			
Class	Top	Left	Right
0	13.00	4.00	9.00
1	15.00	0.00	15.00

Within Node Probabilities			
Class	Top	Left	Right
0	0.464	1.000	0.375
1	0.536	0.000	0.625

Surrogate		Split	Assoc.	Improve.
1 X1	s	0.107	0.250	0.003
2 X2	s	0.011	0.250	0.003

Competitor	Split	Improve.
1 X3	0.027	0.008
2 X6	1.785	0.008
3 X7	4.436	0.006
4 X2	0.052	0.005
5 X4	3.870	0.004


```

*****
*                               *
*      Node 10: X5              *
*      N: 24                   *
*                               *
*****
=====
=      Terminal Node 6          = *      Node 11              *
=      N: 7                    = *      N: 17                *
=====
*****

```

Node 10 was split on X5

A case goes left if $X5 \leq 0.713$

Improvement = 0.013 Complexity Threshold = 0.012

Node	Cases	Wgt	Counts	Cost	Class
10	24		24.00	0.375	1
-6	7		7.00	0.000	1
11	17		17.00	0.471	0

Class	Weighted Counts		
	Top	Left	Right
0	9.00	0.00	9.00
1	15.00	7.00	8.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.375	0.000	0.529
1	0.625	1.000	0.471

	Surrogate		Split	Assoc.	Improve.
1	X3	s	0.004	0.286	.681818E-03
2	X1	r	0.456	0.143	.257831E-03
3	X7	s	1.580	0.143	.358857E-04
4	X4	s	1.194	0.143	.358857E-04
5	X6	s	0.695	0.143	.358857E-04

	Competitor	Split	Improve.
1	X3	0.024	0.011
2	X6	1.810	0.009
3	X7	5.010	0.008
4	X1	0.467	0.006
5	X2	0.099	0.005

```

*****
*                               *
*           Node 11: X2         *
*           N: 17               *
*                               *
*****
*****
*                               *
*           Node 12             *
*           N: 10               *
*                               *
*****
=====
*                               *
*           Node 12             *
*           N: 10               *
*                               *
*****
=====

```

Node 11 was split on X2

A case goes left if $X2 \leq 0.093$

Improvement = 0.012 Complexity Threshold = 0.016

Node	Cases	Wgt	Counts	Cost	Class
11	17		17.00	0.471	0
12	10		10.00	0.300	1
-9	7		7.00	0.143	0

Weighted Counts

Class	Top	Left	Right
0	9.00	3.00	6.00
1	8.00	7.00	1.00

Within Node Probabilities

Class	Top	Left	Right
0	0.529	0.300	0.857
1	0.471	0.700	0.143

Surrogate		Split	Assoc.	Improve.
1 X1	r	0.374	0.286	.729226E-04
2 X3	r	0.006	0.286	0.002
3 X4	r	1.170	0.286	0.005
4 X5	s	1.082	0.286	.178292E-04
5 X6	r	0.535	0.286	0.005

Competitor		Split	Improve.
1 X3		0.022	0.012
2 X1		0.450	0.011
3 X4		1.226	0.007
4 X6		1.723	0.006
5 X7		4.451	0.006

```

*****
*                               *
*           Node 12: X3         *
*           N: 10              *
*                               *
*****
=====
=      Terminal Node 7      = =      Terminal Node 8      =
=           N: 7           = =           N: 3           =
=====

```

Node 12 was split on X3

A case goes left if $X3 \leq 0.023$

Improvement = 0.019 Complexity Threshold = 0.014

Node	Cases	Wgt	Counts	Cost	Class
12	10		10.00	0.300	1
-7	7		7.00	0.000	1
-8	3		3.00	0.000	0

Class	Weighted Counts		
	Top	Left	Right
0	3.00	0.00	3.00
1	7.00	7.00	0.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.300	0.000	1.000
1	0.700	1.000	0.000

	Surrogate		Split	mAssoc.	Improve.
1	X1	s	0.450	0.667	0.011
2	X6	s	3.190	0.667	0.011
3	X2	r	0.052	0.333	0.008
4	X4	r	1.226	0.333	0.005
5	X7	s	4.451	0.333	0.008

	Competitor		Split	Improve.
1	X1		0.450	0.011
2	X6		3.190	0.011
3	X7		4.451	0.008
4	X2		0.052	0.008
5	X4		1.226	0.005

```
*****
*                               *
*           Node 13: X3         *
*           N: 36               *
*                               *
*****
***** =====
*           Node 14             * =      Terminal Node 14      =
*           N: 29               * =              N: 7           =
***** =====
```

Node 13 was split on X3
A case goes left if X3 <= 0.073
Improvement = 0.019 Complexity Threshold = 0.014

Node	Cases	Wgt	Counts	Cost	Class
13	36		36.00	0.222	1
14	29		29.00	0.103	1
-14	7		7.00	0.286	0

Weighted Counts			
Class	Top	Left	Right
0	8.00	3.00	5.00
1	28.00	26.00	2.00

Within Node Probabilities			
Class	Top	Left	Right
0	0.222	0.103	0.714
1	0.778	0.897	0.286

Surrogate		Split	Assoc.	Improve.
1 X1	r	0.136	0.286	.950675E-03
2 X5	r	0.928	0.286	0.001
3 X6	s	6.585	0.286	.950687E-03
4 X2	s	0.327	0.143	.260621E-04
5 X4	r	1.099	0.143	.461757E-03

Competitor	Split	Improve.
1 X1	0.461	0.012
2 X7	17.070	0.012
3 X5	1.190	0.009
4 X4	1.175	0.006
5 X2	0.355	0.006

```

*****
*                               *
*           Node 14: X1         *
*           N: 29               *
*                               *
*****
***** =====
*           Node 15             * = Terminal Node 13      =
*           N: 28               * = N: 1                  =
***** =====

```

Node 14 was split on X1

A case goes left if $X1 \leq 0.464$

Improvement = 0.008 Complexity Threshold = 0.005

Node	Cases	Wgt	Counts	Cost	Class
14	29		29.00	0.103	1
15	28		28.00	0.071	1
-13	1		1.00	0.000	0

Class	Weighted Counts		
	Top	Left	Right
0	3.00	2.00	1.00
1	26.00	26.00	0.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.103	0.071	1.000
1	0.897	0.929	0.000

Surrogate	Split	Assoc.	Improve.
1 X4 r	1.155	1.000	0.008

Competitor	Split	Improve.
1 X4	1.155	0.008
2 X5	0.951	0.008
3 X6	6.510	0.008
4 X7	23.020	0.008
5 X3	0.065	0.002

```

*****
*                      Node 15: X5                      *
*                      N: 28                              *
*****

=====
=      Terminal Node 11      = =      Terminal Node 12      =
=      N: 1                  = =      N: 27                  =
=====

Node 15 was split on X5
A case goes left if X5 <= 0.951
Improvement = 0.008          Complexity Threshold = 0.005

Node      Cases      Wgt Counts      Cost Class
  15         28         28.00         0.071 1
  -11         1          1.00         0.000 0
  -12         27         27.00         0.037 1

                        Weighted Counts
Class      Top      Left      Right
0           2.00      1.00      1.00
1          26.00      0.00     26.00

                        Within Node Probabilities
Class      Top      Left      Right
0          0.071     1.000     0.037
1          0.929     0.000     0.963

Surrogate      Split      Assoc.      Improve.
1 X6           r          6.510      1.000      0.008
2 X7           r          23.020     1.000      0.008

Competitor      Split      Improve.
1 X6            6.510      0.008
2 X7           23.020      0.008
3 X4            2.056      0.003
4 X1            0.328      0.002
5 X2            0.005      0.002

```

```

*****
*                               *
*      Node 16: X5              *
*      N: 62                    *
*                               *
*****
*****
*                               *
*      Node 17                  *
*      N: 60                    *
*                               *
*****
=====
*                               *
*      Node 17                  *
*      N: 60                    *
*                               *
*      =      Terminal Node 19  *
*      =      N: 2              *
*      =                        *
=====

```

Node 16 was split on X5

A case goes left if X5 <= 1.675

Improvement = 0.016 Complexity Threshold = 0.009

Node	Cases	Wgt	Counts	Cost	Class
16	62		62.00	0.081	1
17	60		60.00	0.050	1
-19	2		2.00	0.000	0

Class	Weighted Counts		
	Top	Left	Right
0	5.00	3.00	2.00
1	57.00	57.00	0.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.081	0.050	1.000
1	0.919	0.950	0.000

Competitor	Split	Improve.
1 X3	0.135	0.013
2 X6	11.520	0.013
3 X7	21.838	0.009
4 X2	2.285	0.004
5 X4	1.031	0.002

```

*****
*               Node 17: X3               *
*               N: 60                     *
*****
===== *****
=   Terminal Node 16   = *           Node 18           *
=       N: 1          = *           N: 59           *
===== *****

```

Node 17 was split on X3

A case goes left if $X3 \leq 0.004$

Improvement = 0.008 Complexity Threshold = 0.005

Node	Cases	Wgt	Counts	Cost	Class
17	60		60.00	0.050	1
-16	1		1.00	0.000	0
18	59		59.00	0.034	1

Class	Weighted Counts		
	Top	Left	Right
0	3.00	1.00	2.00
1	57.00	0.00	57.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.050	1.000	0.034
1	0.950	0.000	0.966

Competitor	Split	Improve.
1 X6	11.520	0.008
2 X7	21.838	0.006
3 X4	1.031	0.002
4 X1	0.172	0.001
5 X2	2.422	0.001


```

*****
*               Node 18: X3               *
*               N: 59                     *
*****
=====
=      Terminal Node 17      = =      Terminal Node 18      =
=              N: 58        = =              N: 1            =
=====

```

Node 18 was split on X3

A case goes left if $X3 \leq 0.234$

Improvement = 0.009 Complexity Threshold = 0.005

Node	Cases	Wgt	Counts	Cost	Class
18	59		59.00	0.034	1
-17	58		58.00	0.017	1
-18	1		1.00	0.000	0

Class	Weighted Counts		
	Top	Left	Right
0	2.00	1.00	1.00
1	57.00	57.00	0.00

Class	Within Node Probabilities		
	Top	Left	Right
0	0.034	0.017	1.000
1	0.966	0.983	0.000

Competitor	Split	Improve.
1 X6	11.520	0.008
2 X7	21.838	0.003
3 X4	1.031	0.003
4 X2	2.422	0.001
5 X1	0.366	0.001

DAFTAR GAMBAR

	Halaman
Gambar 2.1	Pohon Klasifikasi Hasil PemilahSimpul Awal ... 19
Gambar 4.1	Jumlah Perusahaan Manufaktur per Sektor Manufaktur 27
Gambar 4.2	Presentase Masing-masing Kelas <i>Financial Distress</i> dan <i>Non-Financial Distress</i> untuk Tahun 2013-2014..... 34
Gambar 4.3	Persentase Masing-masing Kelas <i>Financial Distress</i> dan <i>Non-Financial Distress</i> untuk Tahun 2012-2013 36
Gambar 4.4	Plot <i>Relative Cost</i> dan Banyaknya Simpul Terminal..... 43
Gambar 4.5	Topologi Pohon Klasifikasi Optimal dengan Pemilihan Pemilah Indeks Gini 44
Gambar 4.6	Struktur Pohon Klasifikasi Optimal..... 46

DAFTAR LAMPIRAN

	Halaman
Lampiran A	Data Sebelum SMOTE Tahun 2013-2014 59
Lampiran B	Data Sebelum SMOTE Tahun 2012-2013 59
Lampiran C	Data Sesudah SMOTE Tahun 2013-2014 60
Lampiran D	Data Sesudah SMOTE Tahun 2012-2013 61
Lampiran E	Output Korelasi Untuk Regresi Logistik Biner..... 62
Lampiran F	Output Deteksi Multikolinearitas 62
Lampiran G	Output Pengujian Parameter Serentak dan Parsial 62
Lampiran H	Output Tabel Klasifikasi 63
Lampiran I	Output Variabel Penting Pohon Klasifikasi Maksimum..... 64
Lampiran J	Output Variabel Penting Pohon Klasifikasi Maksimum..... 64
Lampiran K	Output Grafik <i>Relative Cost</i> 64
Lampiran L	Simpul Pohon Klasifikasi..... 64

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pembahasan klasifikasi *financial distress* pada perusahaan manufaktur dapat disimpulkan beberapa hal sebagai berikut.

1. Berdasarkan statistika deskriptif, sektor industri manufaktur terdiri dari 3 sektor yaitu industri dasar dan kimia sebanyak 62 perusahaan, sektor aneka industri sebanyak 38 perusahaan dan sektor industri barang konsumsi sebanyak 32 perusahaan. Kemudian untuk tahun 2013-2014 pada sektor 1 yaitu sektor industri dasar dan kimia terdapat 10% perusahaan manufaktur yang mengalami masalah *financial distress*, 5% pada sektor aneka industri dan 3% pada sektor industri barang konsumsi. Sehingga total terdapat 17% perusahaan industri manufaktur yang mengalami masalah *financial distress* dari total 132 perusahaan. Sedangkan pada tahun 2012-2013, sektor 1 yaitu sektor industri dasar dan kimia terdapat 11% perusahaan manufaktur yang mengalami masalah *financial distress*, 5% pada sektor aneka industri dan 0,7% pada sektor industri barang dan konsumsi. Sehingga total terdapat 16,7% perusahaan industri manufaktur yang mengalami masalah *financial distress* dari total 132 perusahaan.
2. Berdasarkan *preprocessing imbalanced* data untuk data 2013-2014 dilakukan replikasi sebanyak 4 kali agar data menjadi *balance*, sehingga kelas minor yang awalnya berjumlah 23 menjadi 115 data. Kemudian untuk data tahun 2012-2013 dilakukan replikasi sebanyak 4 kali agar data menjadi *balance*, sehingga kelas minor yang awalnya berjumlah 22 menjadi 110 data.
3. Pada model regresi logistik biner di penujian signifikansi parameter diperoleh model terbaik dimana variabel yang

berpengaruh signifikan terhadap kondisi *financial distress* perusahaan manufaktur adalah variabel rasio likuiditas (X_1) dan rasio aktivitas (X_5), dengan ketepatan klasifikasi pada data *traininig* sebesar 84,8%.

4. Metode CART menghasilkan pohon klasifikasi maksimum sama dengan pohon klasifikasi optimum, dengan simpul utamanya adalah variabel rasio solvabilitas (X_4). Dimana karakteristik kelas 0 “perusahaan manufaktur tidak mengalami *financial distress*” adalah rasio aktivitas kurang dari sama dengan 0,599, dan rasio solvabilitas kurang dari sama dengan 1,020 dan untuk kelas 1 “perusahaan manufaktur mengalami *financial distress*” rasio likuiditas kurang dari sama dengan 0,634, rasio profitabilitas kurang dari sama dengan 0,155, rasio aktivitas kurang dari sama dengan 0,599, dan rasio solvabilitas kurang dari sama dengan 1,020. Dengan nilai ketepatan klasifikasi data *traininig* sebesar 96,3%.
5. Hasil perbandingan kedua metode yaitu untuk data *traininig* diperoleh ketepatan klasifikasi analisis regresi logistik sebesar 84,8% dan ketepatan klasifikasi metode CART sebesar 96,3%. Hal ini menunjukkan bahwa metode CART memiliki nilai ketepatan klasifikasi yang lebih besar 11,5% dari metode regresi logistik biner. Tetapi untuk data testing metode regresi logistik biner 2% lebih unggul dari pada metode CART.

5.2 Saran

Saran untuk peneliti selanjutnya, penambahan variabel prediktor dengan skala nominal sangat disarankan jika memakai metode yang sama dengan alasan kemudahan dalam membaca output untuk karakteristik kelas.

DAFTAR PUSTAKA

- Agresti, A. 1990. *Categorical Data Analysis*. New York: John Wiley dan Sons, Inc.
- Altman, E. I. 1968. *Financial ratios, discriminant analysis and the prediction of corporate bankruptcy*, Journal of Finance, Vol. 23, September, pp. 589-609.
- Ardhinta, N.A. 2009. *Manfaat Rasio Keuangan Untuk Memprediksi Kondisi Financial distress Pada Perusahaan Manufaktur Yang Terdaftar Di Bursa Efek Indonesia*. Universitas Muhammadiyah Surakarta, Surakarta.
- Brahmana, R. K. 2007. *Identifying Financial distress Condition in Indonesia Manufacture Industry*. Birmingham Business School, University of Birmingham United Kingdom.
- Breiman L, Friedman J, Olshen R, dan Stone C. 1993. *Classification and Regression Trees*. Chapman Hall : New York – London.
- Chawla N. V, Bowyer K. W, Hall L. O dan Kegelmeyer W. P. 2002. *SMOTE: Synthetic Minority Oversampling Technique*, Journal of Artificial Intelligence Research.
- Hosmer D.W. and Lemeshow S. 1989. *Applied Logistic Regression*. Wiley, New York,
- Ellen E, dan Juniarti. 2013. *Penerapan Good Corporate Governance, Dampaknya Terhadap Prediksi Financial distress Pada Sektor Aneka Industri dan Barang Konsumsi*. Business Accounting Review, Vol.1, No. 2, 2013. Akuntansi Bisnis Universitas Kristen Petra
- Hadi S. dan Anggraini A. 2008. *Pemilihan Prediktor Delisting Terbaik (Perbandingan Antara The Zmijewski Model,*

- The Altman Model, dan The Springate Model*). Jurnal Akuntansi dan Auditing Indonesia.
- Hartoyo, N. T. 2014. *Prediksi Financial Distress Menggunakan Analisis Diskriminan pada Perusahaan Manufaktur yang Terdaftar di BEI*. Jurnal Akuntansi. Fakultas Ekonomi dan Bisnis. Malang.
- Herlina. 2014. *Financial Distress Prediction Using Particle Swarm Optimazation-Support Vector Machine*. Jurusan Teknik Industri, Institut Teknologi Sepuluh Nopember, Surabaya.
- Hofer, C. W. 1980. *Turn Around Strategies*. Journal of Business Strategy 1.
- Khaulasari, Hani. 2016. *Combine Sampling Least Square Support Vector Machine For Multi Class Imbalanced Data Classification*. Jurusan Statistika, Institut Teknologi Sepuluh Nopember, Surabaya.
- Komalasari, Weita B. 2007. *Metode Pohon Regresi Untuk Eksploratori Data Dengan Variabel yang Banyak dan Kompleks*. Jurnal Fungsional Statistisi Pusat Data dan Informasi Pertanian.
- Kompasiana. 2013. http://www.kompasiana.com/virginiafladian/dampak-krisis-global-terhadap-ketenagakerjaan-di-indonesia_55282627f17e61a01d8b4603. Diakses Tanggal 21 November 2015.
- Platt, H. D dan Platt. M.B. 2006. *Comparing Financial distress and Bankruptcy*. Unpublished Thesis, Northeastern University. Boston.
- Sun J. and Li H. 2011. *Dynamic Financial Distress Prediction Using Instance Selection for The Disposal of Concept Drift*, Journal.
- Soesetio, Y. 2008. *Kepemilikan Manajerial dan Institusional, Kebijakan Dividen, Ukuran Perusahaan, Struktur Aktiva dan Profitabilitas Terhadap Kebijakan Hutang*.

- Jurnal Keuangan dan Perbankan, Vol. 12, No. 3, September 2008.
- Whitaker, R. B. 1999. *The Early Stages of Financial distress*. Journal of Economics and Finance.
- Wulan. I. T. 2015. *Klasifikasi Financial Distress Perusahaan Manufaktur yang Terdaftar di BEI Tahun 2009-2013 Menggunakan Regresi Logistik Biner dan SVM*. Jurusan Statistika, Institut Teknologi Sepuluh Nopember. Surabaya.
- Zhahrotun, U. N. 2013. *Model Prediksi Financial Distress Pada Perusahaan Manufaktur Go Public di Indonesia*. Jurusan Teknik Industri, Institut Teknologi Sepuluh Nopember. Surabaya.

BIODATA PENULIS



Nama panjang penulis adalah Muniroh yang biasanya dipanggil Mun oleh teman-teman penulis. Penulis dilahirkan di Surabaya pada tanggal 25 Agustus 1992. Penulis menempuh pendidikan formal di SD YAPITA Surabaya, SMP YAPITA Surabaya, SMAN 20 Surabaya dan pada jenjang perkuliahan lanjut di Diploma III Jurusan Statistika ITS Surabaya

serta Lintas Jalur S1 Statistika ITS Surabaya. Penulis terdaftar sebagai mahasiswa DIII Jurusan Statistika FMIPA ITS angkatan 2011 dengan NRP 1311030014. Beberapa organisasi kemahasiswaan yang pernah diikuti oleh penulis antara lain UKM Cinta Rebana ITS yang aktif dalam departemen Hublu 2013-2014 dan menjadi panitia BCS 2013 sebagai *Instructure Commite (IC)*. Penulis Bercita-cita melanjutkan kuliah hingga tingkat S3 dengan beasiswa penuh dan menjadi dosen. Akhir kata, apabila pembaca memiliki saran, kritik atau diskusi yang berhubungan mengenai tugas akhir atau diskusi mengenai hal lain dapat menghubungi email : muniro.soeka@gmail.com.